

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2004, 05:57:48 ; Search time 7590 Seconds
(without alignments)
11432.510 Million cell updates/sec

Title: US-10-622-516-1

Perfect score: 2002
Sequence: 1 cgcctccagcgcactctcg.....aaaaaaaaaaaaaaaa 2002

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB	ID	Description
1	2002	100.0	2002	6	AR399482	AR399482 Sequence
2	1712	85.5	2068	9	BC000297	BC000297 Homo sapi
3	1711.2	85.5	3195	9	AK095492	AK095492 Homo sapi
4	1370	68.4	1650	9	HUM3H3M	L25798 Homo sapien
5	1300	64.9	1563	9	BT007302	BT007302 Homo sapi
6	1288	64.3	1685	9	HSCOA5	X66435 H.sapiens m
7	1197.4	59.8	3260	10	BC029693	BC029693 Mus muscu
8	1190.2	59.5	3202	10	BC034317	BC034317 Mus muscu
9	1189.6	59.4	2703	10	BC023851	BC023851 Mus muscu
10	1188	59.3	3278	10	BC031363	BC031363 Mus muscu
11	1180.6	59.0	3275	6	AX700129	AX700129 Sequence
12	1180.6	59.0	3275	6	AX827303	AX827303 Sequence
13	1180.6	59.0	3275	10	RNCHMCA	X52625 Rattus norv
14	1103.2	55.1	127832	9	AL356361	AL356361 Human DNA
15	1103.2	55.1	170834	2	AC021792	AC021792 Homo sapi
16	1037.4	51.8	2673	9	AK025736	AK025736 Homo sapi
17	798.2	39.9	1874	5	CHKMGCOAS	M60657 Chicken HMG
18	794.6	39.7	1824	6	I65510	I65510 Sequence 1
19	681.8	34.1	3098	5	BC042929	BC042929 Xenopus l
20	636	31.8	3205	5	BC049456	BC049456 Danio rer
21	476	23.8	28001	6	AR399483	AR399483 Sequence
22	476	23.8	144405	9	AC114947	AC114947 Homo sapi
23	464	23.2	142657	2	AC008773	AC008773 Homo sapi
24	460.4	23.0	565	6	AX320993	AX320993 Sequence
25	436.4	21.8	615	6	AR225558	AR225558 Sequence
26	436.4	21.8	615	6	AX321714	AX321714 Sequence
27	423.4	21.1	1991	10	AK098104	AK098104 Mus muscu
28	421.8	21.1	1994	10	BC024744	BC024744 Mus muscu
29	421.8	21.1	2011	10	BC014714	BC014714 Mus muscu
30	421.8	21.1	2098	9	BC044217	BC044217 Homo sapi
31	420.2	21.0	2058	6	AX337379	AX337379 Sequence
32	420.2	21.0	2058	6	AX411166	AX411166 Sequence
33	420.2	21.0	2058	9	HSHMGCOAS	X83618 H.sapiens m
34	413.2	20.6	1994	6	AX163779	AX163779 Sequence
35	413.2	20.6	1994	10	RATHMGCOA	M33648 Rat mitoch
36	412	20.6	1744	10	HANHMGCO3	L00326 Syrian hams
37	411.4	20.5	1744	4	SSU90884	U90884 Sus scrofa
38	410	20.5	500	6	BD237076	BD237076 Compounds
39	410	20.5	500	6	AR225476	AR225476 Sequence
40	410	20.5	500	6	AX321597	AX321597 Sequence
41	404	20.2	236502	2	AC098186	AC098186 Rattus no
42	404	20.2	259650	2	AC107176	AC107176 Rattus no
43	401	20.0	506	6	BD265188	BD265188 Compounds
44	401	20.0	506	6	AR401174	AR401174 Sequence
45	401	20.0	506	6	AX192620	AX192620 Sequence

ALIGNMENTS

RESULT 1	AR399482	2002 bp	DNA	linear	PAT 18-DEC-2003
LOCUS	AR399482				
DEFINITION	Sequence 1 from patent US 6620608.				
ACCESSION	AR399482				
VERSION	AR399482.1	GI:40141523			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 2002)				
AUTHORS	Gong,F., Yan,C., Di Francesco,V. and Beasley,E.M.				
TITLE	Isolated human synthase proteins				
JOURNAL	Patent: US 6620608-A 1 16-SEP-2003;				
FEATURES	Location/Qualifiers				

source		1. .2002
ORIGIN		/organism="unknown"
		/mol_type="genomic DNA"
Query Match		100.0%; Score 2002; DB 6; Length 2002;
Best Local Similarity		100.0%; Pried. No. 0;
Matches 2002; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
QY	1	CGCCTCCAGCGACTCTCGGCAAGTCCGGGTGGTTGGCGCTATAAAGCTGGT 60
Db	1	CGCCTCCAGCGACTCTCGGCAAGTCCGGGTGGTTGGCGCTATAAAGCTGGT 60
QY	61	AGCGAAGGGAGGCGCGCGGACTGTCCTTTCGTGGCTCACTCCCTTCTCTGCTGCCG 120
Db	61	AGCGAAGGGAGGCGCGCGGACTGTCCTTTCGTGGCTCACTCCCTTCTCTGCTGCCG 120
QY	121	CTCGGTCACGCTTGCTCTTTCACCATGCTGGATCACTTCCCTTGAATGCAGAGCTTGC 180
Db	121	CTCGGTCACGCTTGCTCTTTCACCATGCTGGATCACTTCCCTTGAATGCAGAGCTTGC 180
QY	181	TGGCCAAAAGATGTTGGGAATGTTGCCCTTGAGATCTATTTTCCCTTCAATATGTGAT 240
Db	181	TGGCCAAAAGATGTTGGGAATGTTGCCCTTGAGATCTATTTTCCCTTCAATATGTGAT 240
QY	241	CAAGCAGAGTTGGAAAAATATGATGGTGTAGATGCTGGAAGTAACATTTGGCTTGGGC 300
Db	241	CAAGCAGAGTTGGAAAAATATGATGGTGTAGATGCTGGAAGTAACATTTGGCTTGGGC 300
QY	301	CAGGCCAAGATGGGCTTCTGCACAGATAGAGAAGATATTAACCTCTTTCATGACTGTG 360
Db	301	CAGGCCAAGATGGGCTTCTGCACAGATAGAGAAGATATTAACCTCTTTCATGACTGTG 360
QY	361	GTTCAGAACTTTATGAGAGAAATAACCTTTCCTATGATTGCATTTGGCGGCTGGAAGTT 420
Db	361	GTTCAGAACTTTATGAGAGAAATAACCTTTCCTATGATTGCATTTGGCGGCTGGAAGTT 420
QY	421	GGAACAGAGACAATCATCGACAATACTAAAGTCTGTGAAGACTAATTTGATGCAGCTGTT 480
Db	421	GGAACAGAGACAATCATCGACAATACTAAAGTCTGTGAAGACTAATTTGATGCAGCTGTT 480
QY	481	GAAAGTCTGGGAATAAGATATAGAAAGGAATCGACACAACATAATGCATGCTATGAGGCG 540
Db	481	GAAAGTCTGGGAATAAGATATAGAAAGGAATCGACACAACATAATGCATGCTATGAGGCG 540
QY	541	ACAGCTGCTGTCTTCAATGCTGTAACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGT 600
Db	541	ACAGCTGCTGTCTTCAATGCTGTAACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGT 600
QY	601	GGGACACATATGCAACATGCTCATATGTTTTTACAAGCCTGATATGCTATGTAATTCCT 660
Db	601	GGGACACATATGCAACATGCTCATATGTTTTTACAAGCCTGATATGCTATGTAATTCCT 660
QY	661	ATAGTAGATGGAATACTCTCCATACAGTGTACCTCAGTGCATTAGACCGCTGTAATCT 720
Db	661	ATAGTAGATGGAATACTCTCCATACAGTGTACCTCAGTGCATTAGACCGCTGTAATCT 720
QY	721	GTTACTGCAAAAAAGATCCATGCGCAGTGGCAGAAAAGAGGAAATGATAAAGATTTTACC 780
Db	721	GTTACTGCAAAAAAGATCCATGCGCAGTGGCAGAAAAGAGGAAATGATAAAGATTTTACC 780
QY	781	TTGAATGATTTTGGCTTCATGATCTTTCACTACCACTAATTTGTAACCTGTTCAAGAAATCT 840
Db	781	TTGAATGATTTTGGCTTCATGATCTTTCACTACCACTAATTTGTAACCTGTTCAAGAAATCT 840
QY	841	CTAGCTCGGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAGATAAAAATAGTATC 900
Db	841	CTAGCTCGGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAGATAAAAATAGTATC 900
QY	901	TATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGAAAGACACCTACTTTGATAGAGAT 960
Db	901	TATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGAAAGACACCTACTTTGATAGAGAT 960

QY	961	GTGAGAAAGGCAATTATGAAGGCTAGCTGTGAACCTTTCAGTGCAGAAAACAAAGGCATCT 1020
Db	961	GTGAGAAAGGCAATTATGAAGGCTAGCTGTGAACCTTTCAGTGCAGAAAACAAAGGCATCT 1020
QY	1021	TTACTGTATCAAAATCGAAATATGTACACATCTTCAGTATATGTTCCCTTGCA 1080
Db	1021	TTACTGTATCAAAATCGAAATATGTACACATCTTCAGTATATGTTCCCTTGCA 1080
QY	1081	TCTGTTCTAGCACAGTACTCACTCAGCAATTAGCAGGAAGAAATTGGAGTGTCTTCT 1140
Db	1081	TCTGTTCTAGCACAGTACTCACTCAGCAATTAGCAGGAAGAAATTGGAGTGTCTTCT 1140
QY	1141	TATGTTCTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCG 1200
Db	1141	TATGTTCTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCG 1200
QY	1201	GGGTCTGCTCTTGATAAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTGATTC 1260
Db	1201	GGGTCTGCTCTTGATAAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTGATTC 1260
QY	1261	AGAACTGTTGTGGCACAGATGTCCTGCTGAATAACATGAAGCTCAGAGAGACCCAT 1320
Db	1261	AGAACTGTTGTGGCACAGATGTCCTGCTGAATAACATGAAGCTCAGAGAGACCCAT 1320
QY	1321	CATTGGTCACTATATTTCCCGAGGGTTCATATGATTCACCTCTTTGAAGGAACGTGTAC 1380
Db	1321	CATTGGTCACTATATTTCCCGAGGGTTCATATGATTCACCTCTTTGAAGGAACGTGTAC 1380
QY	1381	TTAGTTAGGTGGAATGAAAAGCACAGAAGACTTAACGCTCGGCGTCCCACTCCAATGAT 1440
Db	1381	TTAGTTAGGTGGAATGAAAAGCACAGAAGACTTAACGCTCGGCGTCCCACTCCAATGAT 1440
QY	1441	GACACTTGTGATGAAGAGTAGAGACTTGTGCAATCAACATAGCAACTGACATATTCCA 1500
Db	1441	GACACTTGTGATGAAGAGTAGAGACTTGTGCAATCAACATAGCAACTGACATATTCCA 1500
QY	1501	AGCCTCGCAAGAAAGTACCAAGACTCCCTGCCACAGCAGACAACTGAAGCAGCTGTC 1560
Db	1501	AGCCTCGCAAGAAAGTACCAAGACTCCCTGCCACAGCAGACAACTGAAGCAGCTGTC 1560
QY	1561	ATTAGTAATGGGGAACATTAAGATTACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGC 1620
Db	1561	ATTAGTAATGGGGAACATTAAGATTACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGC 1620
QY	1621	ATGGGGTGGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGGATTAATTTTAAAGGA 1680
Db	1621	ATGGGGTGGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGGATTAATTTTAAAGGA 1680
QY	1681	TTACATGTTATGTAAATTTTATGTGACTGACATGGAAGCCTGATGACTATCGTGTACTT 1740
Db	1681	TTACATGTTATGTAAATTTTATGTGACTGACATGGAAGCCTGATGACTATCGTGTACTT 1740
QY	1741	GGGAAAGTCTCTTGGCTCTATTTGCTGACATGCTTCTGTGTGCTTGCCCAATGCCAA 1800
Db	1741	GGGAAAGTCTCTTGGCTCTATTTGCTGACATGCTTCTGTGTGCTTGCCCAATGCCAA 1800
QY	1801	ATGTACTCGAATGATGTTAAGGGCTCTGTAAACTTCATACCTCTTGGCCATTGTATG 1860
Db	1801	ATGTACTCGAATGATGTTAAGGGCTCTGTAAACTTCATACCTCTTGGCCATTGTATG 1860
QY	1861	CATGATGTTTGGTTTTTAAACATGTTAATGAATTTGTACTTCTGTCAAGAAAGCA 1920
Db	1861	CATGATGTTTGGTTTTTAAACATGTTAATGAATTTGTACTTCTGTCAAGAAAGCA 1920
QY	1921	GAGTACTAATCTCCAATTAAAAAATTTTAAACATGTAAAAAATTTTAAAAA 1980
Db	1921	GAGTACTAATCTCCAATTAAAAAATTTTAAACATGTAAAAAATTTTAAAAA 1980
QY	1981	AAAAAAAAAAAAAAAAAAAA 2002
Db	1981	AAAAAAAAAAAAAAAAAAAA 2002

RESULT 2	
BC000297	
LOCUS	BC000297
DEFINITION	Homo sapiens 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble), mRNA (cDNA clone IMAGE:2819708), partial cds.
ACCESSION	BC000297
VERSION	BC000297.2 . GI:33991030
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 2068) Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mulahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Vallalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smallus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE	22388257
PUBMED	12477932
REFERENCE	2 (bases 1 to 2068)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	On Aug 20, 2003 this sequence version replaced gi:12653064. Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: DCTD/DTF CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@nhgri.nih.gov
FEATURES	<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov</p> <p>Series: IRAL Plate: 1 Row: a Column: 23</p> <p>This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504428.</p> <p>Location/Qualifiers</p> <p>1..2068</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p>

	gene	/db_xref="taxon:9606" /clone="IMAGE:2819708" /tissue_type="lung, small cell carcinoma" /clone_lib="NIH_MGC_7" /lab_host="DH10B-R" /note="Vector: POTB7" <1..2068 /gene="HMGCS1" /note="synonym: HMGCS" /db_xref="locusid:3157" /db_xref="MIM:142940" <1..1668 /gene="HMGCS1" /codon_start=1 /product="HMGCS1 protein" /protein_id="AAH00297.2" /db_xref="GI:33991031" /db_xref="locusid:3157" /db_xref="MIM:142940" /translacion="HSLSSAARSRLCPKEETVTDLETAVLYPSHSTFMPGSLPLNAEACWPKDVGIVALEIYFSPQYVDQAELEKDGVDAGKYTIGLOAKMGFCTDREDINS LMCVTVQNLMEERNNLSDYDCIGRLVEGTETIIDKSKSVKTNLMQLFEEGNTDIEGIDT TNACTYGVAVENAVNWMISSSWDGRYALVAGDIAYVATGNARPTGAVGAVALLIGP NAPLIFERGLRGMQHAAYDFYKPDMLSEYPIVDGLSIOCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEAFQGVKLEDITYFDRDVEKAFMKASSELFSQTKASLIVSNONGNMYTSVYGSLASVLAQYSPQQLAGRIKIVFSYSGSLATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLREDTHLVNVIPOGSIIDSLFEGTWYLVRVDEKRRTYARPTPNDDTLDEGVGLVHSNIATEHIIPSPAKVPRLPATAAEPEAAVISNGEH" 142..1512 /gene="HMGCS1" /note="HMG CoA synt; Region: Hydroxymethylglutaryl-coenzyme A synthase" /db_xref="CD:pfam01154"
	CDS	
	misc_feature	
ORIGIN		
	Query Match	85.5%; Score 1712; DB 9; Length 2068;
	Best Local Similarity	93.6%; Pred. No. 0;
	Matches 1848; Conservative	0; Mismatches 0; Indels 126; Gaps 1;
QY	134 GCTCTTTCACCATGCGCTGATCATTCTCTTGAATGAGAAGCTTGCGCAAAAGATG	193
DB	95 GCTCTTCACCATGCGCTGATCATTCTCTTGAATGAGAAGCTTGCGCAAAAGATG	154
QY	194 TGGGAATTGTTGCCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGATTGG	253
DB	155 TGGGAATTGTTGCCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGATTGG	214
QY	254 AAAAATATGATGCTGTAGATGCTGGAAGTATACCATTTGGCGCCAGGCCAAGATGG	313
DB	215 AAAAATATGATGCTGTAGATGCTGGAAGTATACCATTTGGCGCCAGGCCAAGATGG	274
QY	314 GCTTCTGCACAGATAGAGAATATTAACTCTCTTGCATGACTGTGTTCAGAAATCTTA	373
DB	275 GCTTCTGCACAGATAGAGAATATTAACTCTCTTGCATGACTGTGTTCAGAAATCTTA	334
QY	374 TGGAGAGAAATAACCTTTCCTATGATTGCATTTGGCGCGCTGGAAGTTGGAACAGACAA	433
DB	335 TGGAGAGAAATAACCTTTCCTATGATTGCATTTGGCGCGCTGGAAGTTGGAACAGACAA	394
QY	434 TCATCGACAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGA	493
DB	395 TCATCGACAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGA	454
QY	494 ATACAGATATGAGAGGAATGCACACAACATAATGCATGCTATGAGGCACAGCTGCTGCT	553
DB	455 ATACAGATATGAGAGGAATGCACACAACATAATGCATGCTATGAGGCACAGCTGCTGCT	514
QY	554 TCAATGCTGTTAAGTGAATGATGATCCAGCTCTTGGAT-----	591
DB	515 TCAATGCTGTTAAGTGAATGATGATCCAGCTCTTGGATGACGGTATGCCCTGTAATTG	574
QY	592 -----	591

Db 575 CAGGAGATATTGCTGTATATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGCAG 634
QY 592 -----GGGCTTCGTGGGACAC 607
Db 635 TAGCTCTGCTAATTGGGCCAAATGCTCTTTAATTTTGAACGAGGGCTTCGTGGGACAC 694
QY 608 ATATGCAACATGCTATGATTTTACAGCCCTGATATGCTATCTGAATATCTATATAGTAG 667
Db 695 ATATGCAACATGCTATGATTTTACAGCCCTGATATGCTATCTGAATATCTATATAGTAG 754
QY 668 ATGAAAACTCTCCATACAGTGTACTCTCAGTGCATTAGACCCTGCTACTCTGTCTACT 727
Db 755 ATGAAAACTCTCCATACAGTGTACTCTCAGTGCATTAGACCCTGCTACTCTGTCTACT 814
QY 728 GCAAAAAAGATCCATGCCCCAGTGGCAGAAAAAGGGAAATGATTAAGATTTTACCTTGAATG 787
Db 815 GCAAAAAAGATCCATGCCCCAGTGGCAGAAAAAGGGAAATGATTAAGATTTTACCTTGAATG 874
QY 788 ATTTGGCTTCATGATCTTTCACACTCACCATATTTGTAACTGGTTCAGAAATCTTAGCTC 847
Db 875 ATTTGGCTTCATGATCTTTCACACTCACCATATTTGTAACTGGTTCAGAAATCTTAGCTC 934
QY 848 GGATGTTGCTGAATGACTTCTTAATGACAGAAATAGATAAAAAATAGTATCTATAGTG 907
Db 935 GGATGTTGCTGAATGACTTCTTAATGACAGAAATAGATAAAAAATAGTATCTATAGTG 994
QY 908 GCCTGGAAGCCTTTGGGGATGTTAAATAGAAAGACACCTACTTTGATAGATGTGAGA 967
Db 995 GCCTGGAAGCCTTTGGGGATGTTAAATAGAAAGACACCTACTTTGATAGATGTGAGA 1054
QY 968 AGGCATTATGAAGGCTAGCTCTGAACCTTTCACTCAGTCAAAAACAAAGGCATCTTTACTTG 1027
Db 1055 AGGCATTATGAAGGCTAGCTCTGAACCTTTCACTCAGTCAAAAACAAAGGCATCTTTACTTG 1114
QY 1028 TATCAAAATCAAAATGGAAATATGTACACATCTTCACTATATGTTCCCTTGCACTGTTC 1087
Db 1115 TATCAAAATCAAAATGGAAATATGTACACATCTTCACTATATGTTCCCTTGCACTGTTC 1174
QY 1088 TAGCACAGTACTCACCCTCAGCAATTAGCAGGGAAGAAATGGAAGTCTTTCTTATGTT 1147
Db 1175 TAGCACAGTACTCACCCTCAGCAATTAGCAGGGAAGAAATGGAAGTCTTTCTTATGTT 1234
QY 1148 CTGTTTGGCTGCCACTCTGTACTCTTTAAAGTCACACAAGATGCTACACGGGGTCTG 1207
Db 1235 CTGTTTGGCTGCCACTCTGTACTCTTTAAAGTCACACAAGATGCTACACGGGGTCTG 1294
QY 1208 CTCTTGATAAATACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1267
Db 1295 CTCTTGATAAATACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1354
QY 1268 GTGTGGCACACAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGACACCCCATCTTTGG 1327
Db 1355 GTGTGGCACACAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGACACCCCATCTTTGG 1414
QY 1328 TCAACTATATTCCTCCAGGGTTCATATAGATTCACTCTTTGAAGGAACTGTGACTTATGTTA 1387
Db 1415 TCAACTATATTCCTCCAGGGTTCATATAGATTCACTCTTTGAAGGAACTGTGACTTATGTTA 1474
QY 1388 GGGTGGATGAAAAACACAGAAAGACTTACGCTGGCGCTCCCACTCCAAATGATGACACTT 1447
Db 1475 GGGTGGATGAAAAACACAGAAAGACTTACGCTGGCGCTCCCACTCCAAATGATGACACTT 1534
QY 1448 TGGATGAAGGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTCCAAGCCCTG 1507
Db 1535 TGGATGAAGGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTCCAAGCCCTG 1594
QY 1508 CCAAGAAAGTACCAAGACTCCCTGACACAGCAGAGAACTGAAAGCAGCTGTCAATTAGTA 1567
Db 1595 CCAAGAAAGTACCAAGACTCCCTGACACAGCAGAGAACTGAAAGCAGCTGTCAATTAGTA 1654
QY 1568 ATGGGGAACATTAAAGATACTCTGTGAGGTGCAAGACTTACAGGGTGGGTGGCATGGGGT 1627

Db 1655 ATGGGGAACATTAAAGATACTCTGTGAGGTGCAAGACTTACAGGGTGGGTGGCATGGGGT 1714
QY 1628 GGGGATATGGAAACAGTTGGAGGAATGGGATATCTGGGGATTAATTTTAAAGATTACATG 1687
Db 1715 GGGGATATGGAAACAGTTGGAGGAATGGGATATCTGGGGATTAATTTTAAAGATTACATG 1774
QY 1688 TTATGTAATTTTATATGTGACTGACATGAGCCCTGGAATGACTATCGTACTTGGGAAG 1747
Db 1775 TTATGTAATTTTATATGTGACTGACATGAGCCCTGGAATGACTATCGTACTTGGGAAG 1834
QY 1748 TCTCTTTGCTCTATTTGCTGACATGCTTCTGTTGTGCTCTGGCCAATGCCAATGACT 1807
Db 1835 TCTCTTTGCTCTATTTGCTGACATGCTTCTGTTGTGCTCTGGCCAATGCCAATGACT 1894
QY 1808 CGAATGATGTTAAGGCTCTGTAAACTTCATACCTCTTTGGCCATTGATGATGATG 1867
Db 1895 CGAATGATGTTAAGGCTCTGTAAACTTCATACCTCTTTGGCCATTGATGATGATG 1954
QY 1868 TTTGTTTTTAAACATGCTATATGAAATTTGTACTTCTGTCAGAAAGACAGAGTAC 1927
Db 1955 TTTGTTTTTAAACATGCTATATGAAATTTGTACTTCTGTCAGAAAGACAGAGTAC 2014
QY 1928 TAATCTCCAATTAAAAATTTTAAACATGTAAAAAATTTTAAACATGTAAAAA 1981
Db 2015 TAATCTCCAATTAAAAATTTTAAACATGTAAAAAATTTTAAACATGTAAAAA 2068
REFERENCE
AUTHORS
1 Oshima,A., Takahashi,Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Ota,T., Wakematsu,A., Ishii,S., Yamamoto,J.,
Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,
Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H.,
Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A.,
Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuhio,Y., Nagai,K.
and Isogai,T.
NEDO human cDNA sequencing project
TITLE
JOURNAL
REFERENCE
AUTHORS
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1. .3195
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="FCBBF1000053"
/tissue_type="brain"
/clone_lib="FCBBF1"
/dev_stage="fetal"

QY	1845	TTTGGCCATTTGATGCATGATGTTGGTTTTTAAACATGGTAATGAATGTGTA	1904
Db	1861	TTTGCCATTTGTGATGCATGATGTTGGTTTTTAAACATGGTAATGAATGTGTA	1920
QY	1905	CTGTCAAGAAAGCAGAGGTACTAATCTCCAAATTAATAAATTTTTTAACATGTA	1964
Db	1921	CTGTCAAGAAAGCAGAGGTACTAATCTCCAAATTAATAAATTTTTTAACATGTA	1980
QY	1965	AAAAAAAAAAAAAAAAAAAAAAAAAAAAA	1994
Db	1981	TTGTACTTTGAACACAAAGATTACAGAAA	2010
RESULT 4			
HUM3H3M		1650 bp	mRNA
LOCUS			linear
DEFINITION		Homo sapiens 3-hydroxy-3-methylglutaryl coenzyme A synthase mRNA,	PRI 24-FEB-1995
ACCESSION		complete cds.	
VERSION		L25798.1	GI:410027
KEYWORDS		3-hydroxy-3-methylglutaryl coenzyme A synthase.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1 (bases 1 to 1650)	
TITLE		Rokosz, L., Boulton, D.A., Butkiewicz, E.A., Sanyal, G., Cuetto, M.A., Lachance, P.A. and Hermes, J.D.	
JOURNAL		Human cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase: expression, purification, and characterization of recombinant wild-type and Cys129 mutant enzymes	
MEDLINE		Arch. Biochem. Biophys. 312 (1), 1-13 (1994)	
PUBMED		94304197	
COMMENT		7913309	
FEATURES		Original	
Source		source text: Homo sapiens fetal adrenal cDNA to mRNA.	
		location/Qualifiers	
		1..1650	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/tissue_type="fetal adrenal"	
		22..1584	
		/codon_start=1	
		/product="3-hydroxy-3-methylglutaryl coenzyme A synthase"	
		/protein_id="AAA62411.1"	
		/db_xref="GI:410028"	
CDS		/translation="MPGSLPLNABACMPKDVGVIALEIYFSPQYVDQAELEKXDVDA SKYITIGLQAKKGFCTREDINSLCMTVVQNMERNNLSDYDCIGRLVETETIIDSK SVKTNIMOLFEEGNTDIEGIDTNACYGGTAAVENAVNWIESSWDGRYALVVAADI AVYATGNARPTGVGAVALLIGNAPLIFERGLRGMQHAVDYKPDMLSEYDIVG KLSIQCYLSALDRCYSVYCKLHAQWQKEGNDKDFLNDGFMI FHSPLYCKLVOKSLA RMLNDFLNDQNRDKNSIYSGLEAFGDVKLEDYVDRDVEKAFKASELFSOKTKAS LTVSNQNGNMVTSVYSGLASVLAQYSPOQLAKRIGVFSYSGSLAATLYSLKVTQDA TPGSALDKITASLCDLSRLDSRGVADVFANMKLRREDTHLVNYIPOGSDISLFE GTWYLVRVDEKHKRRTYARRPTPNDDTLDEGVGLVHSNIATEHIFSPAKKVPRLPATAA EPEAAVISNGEH"	
ORIGIN			
Query Match	68.4%;	Score 1370;	DB 9; Length 1650;
Best Local Similarity	92.0%;	Pred. No. 1e-287;	
Matches 1509;	Conservative 0;	Mismatches 5;	Indels 126; Gaps 1;
QY	134	GCTCTTACCAATGCCTGATCACTTCTTGAATGCAGAAGCTTGCTGGCCAAAAGATG	193
Db	11	GCTCTTACCAATGCCTGATCACTTCTTGAATGCAGAAGCTTGCTGGCCAAAAGATG	70
QY	194	TGGGAATGTGTCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGAGTTGG	253
Db	71	TTGGGATGTGTCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGAGTTGG	130
QY	254	AAAAATATGATGGTGTAGATGCTGAAAAGTATACCATTTGGCTGGGCCAGGCCAAGATGG	313
Db	131	AAAAATATGATGGTGTAGATGCTGGGAAAGTATACCATTTGGCTGGGCCAGGCCAAGATGG	190

QY	314	GCTTCTGCACAGATAGAGAATATTAACTCTCTTTGCATGACTGTGGTTCAGAACTCTTA	373
Db	191	GCTTCTGCACAGATAGAGAATATTAACTCTCTTTGCATGACTGTGGTTCAGAACTCTTA	250
QY	374	TGGAGAGAAATACCTTTCCTATGATTGCATTGGGCGGCTGGAACTTGGACAGACAA	433
Db	251	TGGAGAGAAATACCTTTCCTATGATTGCATTGGGCGGCTGGAACTTGGACAGACAA	310
QY	434	TCATCGACAAATCAAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGA	493
Db	311	TCATCGACAAATCAAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGA	370
QY	494	ATACAGATATAGAAAGATCGACCAACTAATGCATGCTATGAGGCGACAGCTGCTGT	553
Db	371	ATACAGATATAGAAAGATCGACCAACTAATGCATGCTATGAGGCGACAGCTGCTGT	430
QY	554	TCAATGCTGTTAACTGATTGAGTCCAGCTCTTGGGAT-----	591
Db	431	TCAATGCTGTTAACTGATTGAGTCCAGCTCTTGGGATGGAACGCTATGCCCTGTAGTTG	490
QY	592	-----	591
Db	491	CAGAGATATTGCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGATTGGAGCAG	550
QY	592	-----GGCTTCGTGGACAC	607
Db	551	TAGCTCTGCTAATTGGGCCAAATGCTCTTAAATTTTGAACGAGGGCTTCGTGGACAC	610
QY	608	ATATGCAACATGCTGTATGATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGTAG	667
Db	611	ATATGCAACATGCTGTATGATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGTAG	670
QY	668	ATGGAATACTCTCCATACAGTGTCACTCACTGATTCAGTCAACCGCTGCTACTCTACT	727
Db	671	ATGGAATACTCTCCATACAGTGTCACTCACTGATTCAGTCAACCGCTGCTACTCTACT	730
QY	728	GCAAAAGATCCATGCGCCAGTGGCAGAAAGAGGAAATGATAAAGATTTTAACTTGAATG	787
Db	731	GCAAAAGATCCATGCGCCAGTGGCAGAAAGAGGAAATGATAAAGATTTTAACTTGAATG	790
QY	788	ATTTGCTCATGATCTTTCACTCACTCAATATTGTAACCTGGTTCAAGAACTCTAGCTC	847
Db	791	ATTTGCTCATGATCTTTCACTCACTCAATATTGTAACCTGGTTCAAGAACTCTAGCTC	850
QY	848	GGATGTGCTGAATGATCTTCTTAATGACCAATATTGTAACCTGGTTCAAGAACTCTAGCTC	907
Db	851	GGATGTGCTGAATGATCTTCTTAATGACCAATATTGTAACCTGGTTCAAGAACTCTAGCTC	910
QY	908	GCCTGGAAGCCTTTGGGGATGTTAAATTAGAAAGACACCTTCTTGATAGAGATGTGAGA	967
Db	911	GCCTGGAAGCCTTTGGGGATGTTAAATTAGAAAGACACCTTCTTGATAGAGATGTGAGA	970
QY	968	AGGCATTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAACAAAGGCATCTTACTTG	1027
Db	971	AGGCATTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAACAAAGGCATCTTACTTG	1030
QY	1028	TATCAAAATCAAAATGGAATATGTACACATCTTCAATATGTTCCCTTGCACTCTGTT	1087
Db	1031	TATCAAAATCAAAATGGAATATGTACACATCTTCAATATGTTCCCTTGCACTCTGTT	1090
QY	1088	TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAAATGGAATGTTTCTTATGGTT	1147
Db	1091	TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAAATGGAATGTTTCTTATGGTT	1150
QY	1148	CTGGTTTGGCTGCACCTGTACTCTCTTAAAGTACACAAAGATGCTACACCGGGTCTG	1207
Db	1151	CTGGTTTGGCTGCACCTGTACTCTCTTAAAGTACACAAAGATGCTACACCGGGTCTG	1210
QY	1208	CTCTTGATAAATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAAGAACTG	1267
Db	1211	CTCTTGATAAATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAAGAACTG	1270


```

QY      1268 GTGTGGCACCAGATGTCCTTCGCTGTAAGAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1327
      |||||||
Db      1271 GTGTGGCACACGATGTCCTTCGCTGTAAGAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1330

QY      1328 TCAACTATATTTCCCGAGGGTTCAATAGATTCACTCTTTGAAGAAAGTGGTACTTAgTTA 1387
      |||||||
Db      1331 TCAACTATATTTCCCGAGGGTTCAATAGATTCACTCTTTGAAGAAAGTGGTACTTAgTTA 1390

QY      1388 GGGTGGATGAAAGACACAGAAGACTTACGCTCGCGCTCCCACTCCAATGATGACACTT 1447
      |||||||
Db      1391 GGGTGGATGAAAGACACAGAAGACTTACGCTCGCGCTCCCACTCCAATGATGACACTT 1450

QY      1448 TGGATGAAGGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTTCCAAGCCCTG 1507
      |||||||
Db      1451 TGGATGAAGGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTTCCAAGCCCTG 1510

QY      1508 CCAAGAAAGTACCAAGACTCCCTGCCACACAGCAGCAGAACTGGAAGCAGCTGTCAATTAGTA 1567
      |||||||
Db      1511 CCAAGAAAGTACCAAGACTCCCTGCCACACAGCAGCAGAACTGGAAGCAGCTGTCAATTAGTA 1570

QY      1568 ATGGGGAACATTAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTCATGGGCT 1627
      |||||||
Db      1571 ATGGGGAACATTAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTCATGGGCT 1630

QY      1628 GGGGGTATGGGAACAGTTGG 1647
      |||||||
Db      1631 GGGGGTATGGGAACAGTTGG 1650
```

```

RESULT 5
BT007302          1563 bp      mRNA      linear      PRI 13-MAY-2003
LOCUS      Homo sapiens 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
DEFINITION (soluble) mRNA, complete cds.
ACCESSION  BT007302
VERSION     BT007302.1  GI:30583442
KEYWORDS   FLI CDNA.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE  1 (bases 1 to 1563)
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS    Kaline,N., Chen,X., Rolfe,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
TITLE      Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1563)
            Kaline,N., Chen,X., Rolfe,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
TITLE      Direct Submission
JOURNAL    Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT    This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been directionally cloned using BD In-Fusion(TM)
            cloning system between the Sali and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone: 'ACC' after Sali site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
            location/Qualifiers
FEATURES
    source
        1..1563
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="GH00878X1.0"
            /clone_lib="BD Creator(TM) CDS Library derived from MGC
            collection"
```

```

CDS
    /lab_host="DH5alpha TI resistant"
    /note="Vector: pDNR-Dual"
    1..1563
        /codon_start=1
        /product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
        (soluble)"
        /protein_id="AAP35966.1"
        /db_xref="GI:30583443"
        /translation="MPSGLPLNAEACWPKDVGIVALEIFYPSQYVDQAELEKYDGVDA
        GKYYTIGLQAKMGFTCDREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTEITIDSK
        SVKTNLMQLFEESGNTDIEGIDTNAICYGTAFAVENAVNMIESSSWDGRALVAGDI
        AVYATGNARPTGSGAVALLIGPNAPLIFERGLRGTMOHAYDFYKPMLESEPIVDG
        KLSIQCYLSALDRCYSVYCKKIHQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLA
        RMLNDFLNDONRDKNSIYSGLEAFGVDVLEDTYFDRDVEKAFMKASSEFSQTKAS
        LTVSNQNGNMYTSSVYGSGLASVLAQYSPQQLAKRIGVFSYGSGLAATLYSLKVQDA
        TPGSALDKXTASLCDLKSRLDSRTGVAPDVAENMKLREDTHHLVNYIPQGISIDLFE
        GTWYLVVRVDEKHRRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAA
        EPEAAVISNGEH"

ORIGIN
Query Match      64.9%; Score 1300; DB 9; Length 1563;
Best Local Similarity 91.9%; Pred. No. 1.7e-272;
Matches 1436; Conservative 0; Mismatches 0; Indels 126; Gaps 1;

QY      145 ATGCCGTGATCACTTCCTTGAATGCAGAAGCTTGCGCCAAAAGATGTGGGAATTGTT 204
      |||||||
Db      1  ATGCCGTGATCACTTCCTTGAATGCAGAAGCTTGCGCCAAAAGATGTGGGAATTGTT 60

QY      205 GCCCTTGAGATCTATTTTCCTTCTCAATATGTGTGATCAAGACAGAGTTGGAAAAATATGAT 264
      |||||||
Db      61 GCCCTTGAGATCTATTTTCCTTCTCAATATGTGTGATCAAGACAGAGTTGGAAAAATATGAT 120

QY      265 GGTGTAGATGCTGGAAGATATACCATTTGGCTTGGGCCAGGCCAAGATGGGCTTTCGACA 324
      |||||||
Db      121 GGTGTAGATGCTGGAAGATATACCATTTGGCTTGGGCCAGGCCAAGATGGGCTTTCGACA 180

QY      325 GATAGAGAAGATATTAACTCTTTGCATGACTGTGTTCAAGATCTTATGAGAGAAAT 384
      |||||||
Db      181 GATAGAGAAGATATTAACTCTTTGCATGACTGTGTTCAAGATCTTATGAGAGAAAT 240

QY      385 AACCTTTCCTATGATTGCAATTTGGCGGCTGGAAGTTGGAACAGACAATCATGCACAA 444
      |||||||
Db      241 AACCTTTCCTATGATTGCAATTTGGCGGCTGGAAGTTGGAACAGACAATCATGCACAA 300

QY      445 TCAAAGTCTGTGAAGACTTAATTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATTA 504
      |||||||
Db      301 TCAAAGTCTGTGAAGACTTAATTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATTA 360

QY      505 GAAGGAATGCACACAACCTTAATGCATGCTATGAGGCGACAGCTGCTCTTCAATGCTGTT 564
      |||||||
Db      361 GAAGGAATGCACACAACCTTAATGCATGCTATGAGGCGACAGCTGCTCTTCAATGCTGTT 420

QY      565 AACTGATTGAGTCCAGCTCTTGGGAT----- 591
      |||||||
Db      421 AACTGATTGAGTCCAGCTCTTGGGATGGA CGGTATGCCCTGTGATTG CAGAGATATT 480

QY      592 ----- 591
Db      481 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGACAGTAGCTCTGCTA 540

QY      592 -----GGGCTTCGTGGGACACACATATGCACAT 618
      |||||||
Db      541 ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGGCTTCGTGGGACACATATGCACAT 600

QY      619 GCCTATGATTTTTCACAGCCTGATATGCTATCTGAATATCCTATAGTAGTGAAAACTC 678
      |||||||
Db      601 GCCTATGATTTTTCACAGCCTGATATGCTATCTGAATATCCTATAGTAGTGAAAACTC 660

QY      679 TCATACAGTGCTACCTCAGTGCATTAGACCGGTGCTACTCTGTCTACTGCAAAAAGATC 738
      |||||||
Db      661 TCATACAGTGCTACCTCAGTGCATTAGACCGGTGCTACTCTGTCTACTGCAAAAAGATC 720

QY      739 CATGCCCAGTGCAGAAAGAGGAAATGATAAAGATTTTAACTTGAATGATTTTGGCTTC 798
```


Db	721	CATGCCCACTGGCAGAAAAGAGGAATGATAAGATTTTAACTTGATGATTTTGGCTTC	780
QY	799	ATGATCTTTCACTCACCATATTTGTAACCTGGTTCAGAAATCTTAGCTCGATGTGCTG	858
Db	781	ATGATCTTTCACTCACCATATTTGTAACCTGGTTCAGAAATCTTAGCTCGATGTGCTG	840
QY	859	AATGACTTCCCTTAATGACCAAGATAGAGATAAAAAATAGTATCTATAGTGGCCTGGAAGCC	918
Db	841	AATGACTTCCCTTAATGACCAAGATAGAGATAAAAAATAGTATCTATAGTGGCCTGGAAGCC	900
QY	919	TTTGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGAAAGGCATTATG	978
Db	901	TTTGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGAAAGGCATTATG	960
QY	979	AAGGCTAGCTCTGAACCTCTTCAGTCAAGAAAACAAAGGCATCTTACTTGTAATCAATCAA	1038
Db	961	AAGGCTAGCTCTGAACCTCTTCAGTCAAGAAAACAAAGGCATCTTACTTGTAATCAATCAA	1020
QY	1039	AATGAAATATGTACACATCTTCAGTATATGCTCCCTTGCACTGTGTTCTAGCACAGTAC	1098
Db	1021	AATGAAATATGTACACATCTTCAGTATATGCTCCCTTGCACTGTGTTCTAGCACAGTAC	1080
QY	1099	TCACCTCAGCAATTAGCAGGGGAAGAGAAATTGGAGTGTCTTATGCTCTGGTTGGCT	1158
Db	1081	TCACCTCAGCAATTAGCAGGGGAAGAGAAATTGGAGTGTCTTATGCTCTGGTTGGCT	1140
QY	1159	GCCACTCTGTACTCTCTTAAAGTCACACAGATGTACACGGGGGTCTGCTCTTGATAAA	1218
Db	1141	GCCACTCTGTACTCTCTTAAAGTCACACAGATGTACACGGGGGTCTGCTCTTGATAAA	1200
QY	1219	ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTGTGTGGCACCA	1278
Db	1201	ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTGTGTGGCACCA	1260
QY	1279	GATGCTTCGCTGAAAAACATGAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT	1338
Db	1261	GATGCTTCGCTGAAAAACATGAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT	1320
QY	1339	CCCCAGGTTCAATAGATTCACTCTTTGAAAGAAAGTGTACTTAGTTAGGCTGATGAA	1398
Db	1321	CCCCAGGTTCAATAGATTCACTCTTTGAAAGAAAGTGTACTTAGTTAGGCTGATGAA	1380
QY	1399	AAGCACAGAAAGAACTTACGCTCGGCGTCCCACTCCAATGATGACACTTTGGATGAAGA	1458
Db	1381	AAGCACAGAAAGAACTTACGCTCGGCGTCCCACTCCAATGATGACACTTTGGATGAAGA	1440
QY	1459	GTAGACTTGTGCATTCACAATAGCACTGAGCATATTTCCAAAGCCCTGCCAAGAAAGTA	1518
Db	1441	GTAGACTTGTGCATTCACAATAGCACTGAGCATATTTCCAAAGCCCTGCCAAGAAAGTA	1500
QY	1519	CCAAGACTCCCTGCCACAGAGCAGAACTGAAGCAGCTGTCAATTAGTAATGGGAAACAT	1578
Db	1501	CCAAGACTCCCTGCCACAGAGCAGAACTGAAGCAGCTGTCAATTAGTAATGGGAAACAT	1560
QY	1579	TA 1580	
Db	1561	TA 1562	
RESULT 6			
HSCOAS			
LOCUS	HSCOAS	1685 bp	mRNA
DEFINITION	H.sapiens mRNA for HMG-CoA-synthase.		linear
ACCESSION	X66435	S48133	
VERSION	X66435.1	GI:30008	
KEYWORDS	Hydroxymethylglutaryl CoA Synthase.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Russ,A.P., Ruzicka,V., Maerz,W., Appelhans,H. and Gross,W.		

TITLE	Amplification and direct sequencing of a cDNA encoding human cytosolic 3-hydroxy-3-methylglutaryl-coenzyme A synthase			
JOURNAL	Biochim. Biophys. Acta 1132 (3), 329-331 (1992)			
MEDLINE	93041939			
PUBMED	1358203			
REFERENCE	2 (bases 1 to 1679)			
AUTHORS	Russ, A.			
TITLE	Direct Submission			
JOURNAL	Submitted (26-MAY-1992) A. Russ, Labor fur angewandte Biochemie, Theodor Stern-Kai 7, W-6000 Frankfurt am Main 70, FRG			
REMARK	revised by [2]			
REFERENCE	3 (bases 1 to 1685)			
AUTHORS	Russ, A.			
TITLE	Direct Submission			
JOURNAL	Submitted (10-AUG-1992) Andreas Russ, Zentrum der biologischen Chemie, J.W.-Goethe-Universitaet Frankfurt, Theodor-Stern-Kai 7, Frankfurt, 6000, Germany			
FEATURES	Location/Qualifiers			
source	1..1685			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="HMGCoASynthase"			
	/cell_type="Fibroblast"			
	123..1685			
	/codon_start=1			
	/product="Hydroxymethylglutaryl CoA Synthase"			
	/protein_id="CAA47061.1"			
	/db_xref="GI:30009"			
	/db_xref="GOA:Q01581"			
	/db_xref="SWISS-PROT:Q01581"			
	/translation="MPGSLPLNAAECWPKDVGVIVALEIFYPSQYVDQAELEKYDVDA GKVTIGLGAKMGFCTDRDINSLSMTVQNLNERNLSYDCIGLEVGTEITIDKSR SVKTNLMQLFEESGNTDIEGIDTTNACYGCTAAVENAVNMIESSWDGRYALVVGDI AVYATGNARPTGVGAVALLIGNAPLIFERGLRGTMOHAYDFYKPDMLSEYPIVDO KLSIQCYLSALDRCYSVYCKKIHQWQKEANDNEFTLNDFFGFMIFHSPYCKLVOKSLA RMLNDFLNDQNRDKNSIYSGKAFGDKLEDTYFDRVEKAFMKASSELFSQKTKAS LLSVNONGNMVTSVYGSLSVLAQYSPQHLAKRIGVFSYGSAAATLXSLKTQDA TPQSALDKITASLDCDLKSRIDSRIGVAQDVFAENMKLREDTHLVNYIPQGSIDSLFE GTWYLVRVDEKHRRTYARPRTPNDDTLDEGVGLVHSNIATEHIPSAPKVPRLPATPA BPEAAVISNGVW"			
ORIGIN				
Query Match	64.3%	Score 1288;	DB 9;	Length 1685;
Best Local Similarity	91.0%;	Pred. No. 7.1e-270;		
Matches 1433;	Conservative 0;	Mismatches 15;	Indels 126;	Gaps 1
QY	134	GCTCTTTCACCATGCGCTGGATCACTTCTTGAATGACAGAACTTGCTGGCCAAAGATG	193	
Db	112	GCTCTTTCACCATGCGCTGGATCACTTCTTGAATGACAGAACTTGCTGGCCAAAGATG	171	
QY	194	TGGGAATGTGGCCCTTGAGATCTATTTCTTCTCAATATGTGATCAAGCAGATTGG	253	
Db	172	TGGGAATGTGGCCCTTGAGATCTATTTCTTCTCAATATGTGATCAAGCAGATTGG	231	
QY	254	AAAAATATGATGCTGTAGATCTGGAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG	313	
Db	232	AAAAATATGATGCTGTAGATCTGGAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG	291	
QY	314	GCTTCTGCACAGATAGAGAGATATTAATCTCTTCTTGCATGACTGTGTTCAGAACTTTA	373	
Db	292	GCTTCTGCACAGATAGAGAGATATTAATCTCTTCTTGCATGACTGTGTTCAGAACTTTA	351	
QY	374	TGGAGAGAAATAACCTTCTATGATTTGCATTTGGCGGCTGGAAGTTGGAACAGAGACAA	433	
Db	352	TGGAGAGAAATAACCTTCTATGATTTGCATTTGGCGGCTGGAAGTTGGAACAGAGACAA	411	
QY	434	TCATCGACAATCAAAGTCTGTGAAGACTAAATTGATGACAGCTGTTGAAGAGTCTGGGA	493	
Db	412	TCATCGACAATCAAAGTCTGTGAAGACTAAATTGATGACAGCTGTTGAAGAGTCTGGGA	471	
QY	494	ATACAGATATGAAGAAATCGACACAATAATGATGCTATAGAGGCACAGCTGCTGCT	553	

Db	472	ATACAGATATAGAAGGAATCGACACAAC	TAATGCATGCTATGAGG	CCACAGCTGCTGCT	531
Qy	554	TCAATGCTGTTAACTGGATTGAGTCCAG	CTCTTGGAT	-----	591
Db	532	TCAATGCTGTTAACTGGATTGAGTCCAG	CTCTTGGATGAGCGGTATGCCCTGTA	GTG	591
Qy	592	-----	-----	-----	591
Db	592	CAGGAGATATTGCTGTATATGCCACA	GAAATGCTAGACCTACAGGTGAGT	GCAGC	651
Qy	592	-----	-----GGGCTTCGTGGACAC	607	
Db	652	TAGCTCTGCTAATTGGGCCAAATGCT	CTTTAATTTTGAACGAGGGCTTCGT	GGACAC	711
Qy	608	ATATGCAACATGCGCTATGATTTTACA	AGCCTGATATGCTATCTGAATATCCT	ATATAGT	667
Db	712	ATATGCAACATGCGCTATGATTTTACA	AGCCTGATATGCTATCTGAATATCCT	ATATAGT	771
Qy	668	ATGAAAACTCTCCATACAGTGTACCT	CAGTGCATTAGACCGCTGCTACTCT	GCTACT	727
Db	772	ATGAAAACTCTCCATACAGTGTACCT	CAGTGCATTAGACCGCTGCTACTCT	GCTACT	831
Qy	728	GCAAAAAGATCCATGCGCCAGTGGC	AGAAAGGGAATGATTAAGATTTTACC	TTGAATG	787
Db	832	GCAAAAAGATCCATGCGCCAGTGGC	AGAAAGGGAATGATTAAGATTTTACC	TTGAATG	891
Qy	788	ATTTGGCTTCATGATCTTCACTCA	CCATATTTGTAACCTGGTTCAGAA	TCTCTAG	847
Db	892	ATTTGGCTTCATGATCTTCACTCA	CCATATTTGTAACCTGGTTCAGAA	TCTCTAG	951
Qy	848	GGATGTTGCTGAATGACTTCCCTTA	TATGACCAGAATAGAGATTAATA	TAGTATCT	907
Db	952	GGATGTTGCTGAATGACTTCCCTTA	TATGACCAGAATAGAGATTAATA	TAGTATAGT	1011
Qy	908	GCTTGAAGCCTTTGGGAGTGTAA	ATTAGAAAGACACCCTACTTTGAT	AGAGATGT	967
Db	1012	GCTTGAAGCCTTTGGGAGTGTAA	ATTAGAAAGACACCCTACTTTGAT	AGAGATGT	1071
Qy	968	AGGCAATTATGAAGGCTAGCTCT	GAACTCTTCAGTCAGAAAA	CAAGGCATCTT	1027
Db	1072	AGGCAATTATGAAGGCTAGCTCT	GAACTCTTCAGTCAGAAAA	CAAGGCATCTT	1131
Qy	1028	TATCAAAATCAAAATGGAATAT	GTACACATCTTCAGTATATG	TTCCCTGCAT	1087
Db	1132	TATCAAAATCAAAATGGAATAT	GTACACATCTTCAGTATATG	TTCCCTGCAT	1191
Qy	1088	TAGCACAGTACTCACCCTCAG	CAATTAGCAGGGAAGAA	TGGAGT	1147
Db	1192	TAGCACAGTACTCACCCTCAG	CAATTAGCAGGGAAGAA	TGGAGT	1251
Qy	1148	CTGTTTGGCTGCGCACTCT	GTACTCTTTAAAGTCACACA	GAAGTCTAC	1207
Db	1252	CTGTTTGGCTGCGCACTCT	GTACTCTTTAAAGTCACACA	GAAGTCTAC	1311
Qy	1208	CTCTGATTAATAATACAGCA	AGTTTAATGTGATCTTAA	TCAAGGCTGAT	1267
Db	1312	CTCTGATTAATAATACAGCA	AGTTTAATGTGATCTTAA	TCAAGGCTGAT	1371
Qy	1268	GTTGGACCAAGATGTCTT	CGCTGAAAAATGAAGCTC	AGAGACACCCAT	1327
Db	1372	GTTGGACCAAGATGTCTT	CGCTGAAAAATGAAGCTC	AGAGACACCCAT	1431
Qy	1328	TCAACTATATTTCCCA	GGGTTCAATAGATTCACT	CTTTGAAGAA	1387
Db	1432	TCAACTATATTTCCCA	GGGTTCAATAGATTCACT	CTTTGAAGAA	1491
Qy	1388	GGGTGATGAAAAAGCAC	AGAAGAACTTACGCTCG	GGCTCCCAAT	1447
Db	1492	GGGTGATGAAAAAGCAC	AGAAGAACTTACGCTCG	GGCTCCCAAT	1551
Qy	1448	TGGATGAGAGTAGGACT	TGTGCAATCAACATAG	CAACTGAGCAT	1507
Db	1552	TGGATGAGAGTAGGACT	TGTGCAATCAACATAG	CAACTGAGCAT	1611

Qy	1508	CCAAGAAAGTACCAGAGACTCCCTG	CCACAGCAGACAGAACTGAAGCAG	CTGTCTATAGTA	1567
Db	1612	CCAAGAAAGTACCAGAGACTCCCTG	CTACACAGCAGACAGAACTGAAGCAG	CTATATAGTA	1671
Qy	1568	ATGGGGAACATTAA	1581		
Db	1672	ATGGGGTATGTAA	1685		
RESULT 7					
BC029693					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REMARK					
COMMENT					

3260 bp
mRNA
linear
ROD 07-OCT-2003
Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
(CDNA clone MGC:36525 IMAGE:5375374), complete cds.
BC029693
BC029693
GI:20988708
MGC.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3260)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Boufard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
human and mouse cDNA sequences
22388257
12477932
2 (bases 1 to 3260)
Strausberg,R.
Direct Submission
Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LNL)
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAK Plate: 53 Row: p Column: 14
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22122370.

Qy	1501	AGCCCTGCCAAGAAAGTACCAGAAGCTCCCTGCCACAGCAGACAAGCTTGAAGCAGCTGTC	1560
Db	1561	AGCCCTGCTAAGAAAGTCCCAAGACTCCCTGCAACCTCGGCCGAATCTGAATCAGCTGTC	1620
Qy	1561	ATTAGTAATGGGGAACATTAAAGTACTCTGTGAGGTGCAGAAGCTTCAGGGTGGGCTGGGC	1620
Db	1621	ATCAGTAACGGGGAGCAGCTGAGAGTCTGTGGCCTTACA-----GAGGCTCGGG	1669
Qy	1621	ATGGGCTGGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGGAATAATTTAAGGA	1680
Db	1670	GCCGGGATGGGGTATGGAACCGTTGGAGGAAT-GGATGTCCTTGGGACAAATTTTGCA-GA	1727
Qy	1681	TTACATGTTATGTAATTTTATGTGACTGACATGAGCCCTGGATGACTATCTGTACTT	1740
Db	1728	TTATGTGTGCTTAATAATGTAATGTAACGACACAGAGCCCAAGAACTATTGTGTTTT	1787
Qy	1741	GCGAAGTCT---CTTGTCTATTGCTGACATGCTTCTGTTGTGCTGCGCAATGC	1797
Db	1788	GGAAAGTCTCTGCTGAACTGAACTGTAACATGCTTCCGTGTGTGCTGGCCAGTGG	1847
Qy	1798	CAATGTACTCGAATGATGTT-AAGGCTCTGTAAACTTCATACCTCTTTGGCAATTG	1856
Db	1848	TAAATGTACTGCAGTGAATGTTAAAGGCTCTGTAGAACTTATACCTCTCTGGATGTTA	1907
Qy	1857	TATGATGATGTTTGTGTTTAAACATGSTATATGAATTGTGACTTCTGTCAGAAGAA	1916
Db	1908	TACGCAATGCAAGTTAGTTTTCAGATGTGTT--ATGAAGTGAATGCTTCTGACAGCAAAA	1964
Qy	1917	AGCAGAGTACTAATCTCCAAATTAAAAA	1945
Db	1965	GGCAGAGTACTAGTGTCCAAATTTTAAA	1993

RESULT 8	BC034317	3202 bp	mRNA	linear	ROD 12-NOV-2003
LOCUS	Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA				
DEFINITION	(CDNA clone MGC:36662 IMAGE:5366786), complete cds.				
ACCESSION	BC034317				
VERSION	BC034317.1	GI:21706865			
KEYWORDS	MGC.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 3202)				
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stajdron,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.				
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)				
MEDLINE	22388257				
PUBMED	12477932				
REFERENCE	2 (bases 1 to 3202)				
AUTHORS	Strausberg,R.				
TITLE	Direct Submission				

JOURNAL	Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk Email: gcgabs-r@mail.nih.gov Tissue Procurement: The Cepko Laboratory cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amg@bcm.tmc.edu Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegé, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.				
FEATURES	<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov</p> <p>Series: IRAK Plate: 53 Row: k Column: 24</p> <p>This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22122370.</p> <p>Location/Qualifiers</p> <p>1..3202</p> <p>/organism="Mus musculus"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="MGC:36662 IMAGE:5366786"</p> <p>/tissue_type="Eye, retina, mouse strain C57BL/6"</p> <p>/clone_lib="NIH MGC_94"</p> <p>/lab_host="DH10B"</p> <p>/note="Vector: pCMV-SPORT6"</p> <p>1..3202</p> <p>/gene="Hmgcs1"</p> <p>/note="synonym: MGC36662"</p> <p>/db_xref="LocusID:208715"</p> <p>/db_xref="MGI:107592"</p> <p>25..1587</p> <p>/codon_start=1</p> <p>/product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1"</p> <p>/protein_id="AAH34317.1"</p> <p>/db_xref="GI:21706866"</p> <p>/db_xref="LocusID:208715"</p> <p>/translation="MPGSLPLNAEACWPKDVGIVALEIYPSQYVDQAELEKYDGVDA GKYTIGLGOARMGFCTDREDINDISLCTVVOKLMERHSLSYDCIGRLFVGTEITIDSK SVKSNLMQLFEESGNTDIEGIDTNNACYGGTAAVFNAVNVWDESSWDRYALVVAAGDI AIVATGNARPTGCVGAVALLIGPNAPLIFDRLRGTHMQHAYDFYKPKDMLSEVYVDG KLSIQCYLSALDRCTSVYRKIRAOQOKEGKDFLINDFGFMI FHSPYCKLVOKSLA RMFLNDFLNDQNRDKNSIYSGLEAFGVDKLEDYFDRDVEKAFMKASSELFNQXTKAS LTVSNONGNMVTSVYGSVLASVLAQSPQOLAGRVGVFSYSGLAATLYSLVTODA TPGSALDKITASLCDLKSRLDSRTCVAIPDVFAENMKLRBDTHLANIIPQCSIDSLFE GTWYLVLRVDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATETHIIPSPAKKVPRLPATSA ESESAVISNGEH"</p> <p>61..1431</p> <p>/note="HMG CoA synt; Region: Hydroxymethylglutaryl-coenzyme A synthase"</p> <p>/db_xref="CDI:pfam01154"</p>				
CDS					
gene					
misc_feature					
ORIGIN					
Query Match	59.5%; Score 1190.2; DB 10; Length 3202;				
Best Local Similarity	79.8%; Pred. No. 1.4e-248;				
Matches 1555; Conservative	0; Mismatches 248; Indels 146; Gaps 7;				
Qy	127	CACGCTGCTCTTTCACCATGCTGGATCAGTCTCTTGAATGCAGAAGCTTGCGCA	186		
Db	7	CCCTTGGCTCTTTCACCATGCTGGGTCACTTCTTGAATGCAGAGGCTTGCGCA	66		
Qy	187	AAAGATGTGGGAATGTTGCCCTTGAGATCTAATTTCTCTCAATATGTTGATCAAGCA	246		
Db	67	AAAGATGTGGGAATGTTGCCCTTGAAATCTACTTCTCTCAATATGTGATCAAGCT	126		

QY	1201	GGGFTGCTCTTGATTAATAACAGCAAGTTTATGTGATCTTTAATCAAGCTTGATTCA	1260
	1207	GGATCTGCCCTTGATTAATAACAGCAAGTTTATGTGACCTTAATCAAGCTTGACTCG <th>1266</th>	1266
QY	1261	AGAACTGGTGTGGCACACAGATGTCTTCGCTGAATAACATGAAGCTCAGAGAGACACCCAT <th>1320</th>	1320
	1267	AGAACTGTGTGGCACCGGATGTCTTTGGCTGAATAACATGAAGCTCAGAGAGACACACAT <th>1326</th>	1326
QY	1321	CATTGGTCAACTATATATCCCCAGGGTTCAATAGATTCACTCTTTGAAGAACTGTGTA <th>1380</th>	1380
	1327	CACCTAGCCCAACTATATATCCCCAGTGTTCATAGACTCACTCTTTGAAGAACTGTGTA <th>1386</th>	1386
QY	1381	TTAGTTAGGGGTGGATGAATAAGCAGAGAATTACGCTCGGCGTCCCACTCCAATGAT <th>1440</th>	1440
	1387	CTGTTCAGAGTGGATGAATAAACACAGAAAGACTTACGCCCGGCCCTTCACAAATGAC <th>1446</th>	1446
QY	1441	GACACTTTGGATGAAGGAGTAGGACTTGTGCATTCAACATAAGCACTGAGCATATTCCA <th>1500</th>	1500
	1447	CACAGTTTGGATGAAGGATGGGGCTCGTGCATAGTAACACAGACACAGACATATTCCA <th>1506</th>	1506
QY	1501	AGCCCTGCCAAGAAGTACCAAGACTCCCTGCGCACAGCAGACAACCTGAAGCAGCTGC <th>1560</th>	1560
	1507	AGCCCTGCTAAGAAAGTCCCAAGACTCCCTGCAACTCGGCCGAATCTGAATCAGCTGTC <th>1566</th>	1566
QY	1561	ATTAGTAATGGGGAACATTAAGATACCTGTGAGGTGCAAGACTTCAGGGTGGGTGGC <th>1620</th>	1620
	1567	ATCAGTAACGGGGAGCAGCTGAGAGTCTGTGGCTTTACA-----GAGGCTCGGG <th>1615</th>	1615
QY	1621	ATGGGCTGGGGGTATGGGACAGCTTGAGAGAACTGGGATATCTGGGGATAATTTAAAGGA <th>1680</th>	1680
	1616	GCCGGGATGGGTATGGAACCGGTTGGAGGAAT-GGATGCTTTGGGACAAATTTGCA-GA <th>1673</th>	1673
QY	1681	TTACATGTTATGTAAATTTTATGTGACTGCATGAGGCTGGAATGACTATCGTACTT <th>1740</th>	1740
	1674	TTATGTGTGCTTAAATGTAAATGTAACTGACACAGAGCCAGAAAGCTATTGTGTTT <th>1733</th>	1733
QY	1741	GGGAAGTCT--CTTGTCTATTGTGTCGATGCTTCCGTGTGTGGTCTGGCCAATGC <th>1797</th>	1797
	1734	GGAAAGTCTCTGCTGAAGTGAAGTGTGTAATGCTTCCGTGTGTGGTCTGGCCAATGC <th>1793</th>	1793
QY	1798	CAATGTACTCGAATGATGTT-AAGGCTCTGTAAAACTTCATACCTTTGGCCATTG <th>1856</th>	1856
	1794	TAAATGTACTGCAATGATGTTAAAGGCTCTGTGAACCTTTATACCTCTCTGATGTTA <th>1853</th>	1853
QY	1857	TATGATGATGTTTGGTTTAAACATGGTATTAATGAATTGTGTACTTCTGTGAGAAGAA <th>1916</th>	1916
	1854	TACGATGCAATGTTAGTTTTCAGATGTGTT--ATGAACGTAGTCTTCTGACAGCAAAA <th>1910</th>	1910
QY	1917	AGCAGAGCTACTAATCTCCAATTTAAAAA <th>1945</th>	1945
	1911	GGCAGAGCTACTAGTGTCCAATTTTAAA <th>1939</th>	1939
Db			
RESULT 9			
BC023851	2703 bp	mRNA	linear
LOCUS			
DEFINITION	Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA		
ACCESSION	BC023851		
VERSION	BC023851.1	GI:23271493	
KEYWORDS	MGC.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE			
AUTHORS	Klausner,R.D., Collins,F.S., Wagner,L., Shermen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buelow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Martusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,		

Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL
MEDLINE
PUBMED
22388257
12477932
2 (bases 1 to 2703)
Strauberg,R.
Direct Submission
Submitted (05-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REFERENCE
AUTHORS
JOURNAL

REMARK
COMMENT
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaapi,R.,
Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantiriop,S., Thomas,P.J., Touchman,J.W.,
Tsurgueon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

FEATURES
source
1. 2703
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:36620 IMAGE:5347038"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
1. 2703
/gene="Hmgcs1"
/note="synonym: MGC36662"
/db_xref="locusID:208715"
/db_xref="MGI:107592"
149. 1711
/codon_start=1
/product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase
1"
/protein_id="AAH23851.1"
/db_xref="GI:23271494"
/db_xref="locusID:208715"
/translation="MPSGLPLNAEACWPKDVGIVALIETFPSSQYVDQAELEKYGVDVA

KYTIIGLQARMGEFCTDREDINSLCLTVQKLMERHSLSYDCIGRLVEVTEITIDSK
SVKSNIMQLFEESGNTDIEIDITNACYGGTAVERNVAWNVSSWDGRYALVYAGDI
AIATGNARPTGGVAGVALIKKIRANPILFDKRGTHMOHAYDFYKPDMLSEYPVVG
KLSTQCYLSALDRCSVYRKRIRAQWQKEGKDFLNDQFGFMI FHS PYCKLVQKSLA
RMFLNDFLNDQNRDKNISYSGLEAFGDKLEDYFDRDVEKAFMKASSELFNQTKAS
LLVSNQNGNMVTSYVYGLASVLAQYSPQOLAGKRVGVFSYSGSLATLYSLKTYQDA
TPGSALDKITASTLCSRLDSRTCVAPDVFAENMKLRDTHLANIYIPQCSIDSLFE
GTWYLVRVDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATEHI PSPA KVRPLPATSA
ESESAVISNGEH"
185. 1555
/note="HMG CoA synt; Region:
Hydroxymethylglutaryl-coenzyme A synthase"
/db_xref="CDD:pfam01154"

ORIGIN
Query Match 59.4%; Score 1189.6; DB 10; Length 2703;
Best Local Similarity 79.9%; Pred. No. 1.9e-248;
Matches 1552; Conservative 0; Mismatches 244; Indels 146; Gaps 7;

QY	134	GCTCTTCAACCATGCGCTGATCACTTCTTGAATGCAGAACCTTGCGCCAAAGATG	193
DB	138	GCTCTTCAACCATGCGCTGATCACTTCTTGAATGCAGAACCTTGCGCCAAAGATG	197
QY	194	TGGGAATTGTTGCCCTTGAGATCTATTCTCTCTCAATATGTTGATCAAGCAGATTGG	253
DB	198	TGGGAATGTTGCCCTTGAGATCTATTCTCTCTCAATATGTTGATCAAGCAGATTGG	257
QY	254	AAAAATATGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTG	313
DB	258	AAAAATATGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTG	317
QY	314	GCTTCTGACAGATAGAGAAGATATTAATCTCTTGTGATGATGATGATGATGATGATG	373
DB	318	GCTTCTGACAGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATG	377
QY	374	TGAGAGAATAAATACCTTCTCTGATGATGATGATGATGATGATGATGATGATGATG	433
DB	378	TGAGAGAATAAATACCTTCTCTGATGATGATGATGATGATGATGATGATGATGATG	437
QY	434	TCAATGACAAATCAAGTCTGTGAAGACTAATTTGATGATGATGATGATGATGATGATG	493
DB	438	TCAATGACAAATCAAGTCTGTGAAGACTAATTTGATGATGATGATGATGATGATGATG	497
QY	494	ATACAGATATAGAAGAAATCGACAACTAATGATGATGATGATGATGATGATGATGATG	553
DB	498	ATACAGATATAGAAGAAATCGACAACTAATGATGATGATGATGATGATGATGATGATG	557
QY	554	TCAATGCTGTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	591
DB	558	TCAATGCTGTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	595
QY	592	-----	591
DB	618	CAGAGACATTGCTATATATGCCACAGAAATGCCAGACTACAGGTGAGCTG	677
QY	592	-----GGGCTTCGTGGACAC	607
DB	678	TGGCCCTGCTAATTGGGCCAAACGCTCTTAATTTTGGACCGAGGGCTCCGTGGACAC	737
QY	608	ATATGACATGCTGATGATTTTACAAGCCTGATATGCTATCTGAATATCTATAGTAG	667
DB	738	ACATGACATGCTGATGATTTTACAAGCCTGATATGCTATCTGAATATCTATAGTAG	797
QY	668	ATGAAAATCTCTCCATACAGTGTCACTCAGTGATTAAGCCGCTGCTACTGTCTACT	727
DB	798	ACGGAAGCTCTCCATACAGTGTCACTCAGTGATTAAGCCGCTGCTACTGTCTACT	857
QY	728	GCAAAAGATCCATGCGCAGTGGGAGAAAGGAAATGATTAAGATTTTACCTTGAATG	787
DB	858	GCAAAAGATCCATGCGCAGTGGGAGAAAGGAAATGATTAAGATTTTACCTTGAATG	917
QY	788	ATTTGGCTTCATGATCTTTCATCAACCATATTTGTAACCTGGTTCAAGAAATCTAGCTC	847

Db 918 ATTTTGCTTCANGATCTTTCACTCACCATATTGTAAACGTGTCAGAAATCTCTAGCTC 977
QY 848 GGATGTTGCTGAATGACTTCTCTAAATGACACAGAAATAGATMAAAATAGTATCTATAGTG 907
Db 978 GGATGTTCTGAATGACTTCTCTAAATGATCAAAACAGAGATMAAAACAGTATTTACAGTG 1037
QY 908 GCCTGGAAAGCCTTGGGGATGTTAAATTAGAAAGACCTACTTTGATAGAGATGGAGAGA 967
Db 1038 GACTGGAAGCCTTGGGGACGTTAAATTAGAAAGATCTTACTTTGACAGAGATGAGAAA 1097
QY 968 AGGCATTTATGAAGGCTAGCTCTGAACCTCTCAGTCAGAAACAAAGGCATCTTTACTTG 1027
Db 1098 AGGCATTTATGAAGGCTAGTCTGAGCTATTCAACAGAAACAAAGCGCTTTGCTTG 1157
QY 1028 TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTGCCCTTGCACTGTTTC 1087
Db 1158 TGTCTAATCAGAAATGGAATATGTACACATCCTGTCTATGTTCCCTGCTTCTGTCC 1217
QY 1088 TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAAATTGAGTGTCTTTCTTATGTTT 1147
Db 1218 TGGCACAGTACTCACTCAGCAGTTGGCAGGGAAGGGGTTGAGTGTCTCTTACGGTT 1277
QY 1148 CTGCTTTGGCTGCGCACTCTGTACTCTTTAAAGTCAACAAGATGCTACACCGGGGCTG 1207
Db 1278 CTGGCTTGGCTGCGCAGCTGTACTCCCTTAAAGTCAACAAGACGCGCACACAGAGATCTG 1337
QY 1208 CTCTTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1267
Db 1338 CCCTTGATAAATTAACAGCAAGTTTATGTGACCTTAAATCAAGGCTTGACTCGAGAACTT 1397
QY 1268 GTGTGGCACCAAGATGTCCTCGCTGAAACATGAAAGCTCAGAGAGACACCCATCATTTGG 1327
Db 1398 GTGTGGCACCCGAGTGTCTTGTGCTGAACATGAAAGCTCAGAGAGACACACATCATTTAG 1457
QY 1328 TCAACTATATTCGCCAGGGTTCATATAGATTCACTCTTTGAAGAACGTGTACTTAGTTA 1387
Db 1458 CCAACTATATTCGCCAGTGTCAATAGACTCACTCTTTGAAGAACGTGTATCTGTCA 1517
QY 1388 GGGTGGATGAAAGACAGAGAAGACTTACGCTCGGCGTCCCACTCCAATGATGACACTT 1447
Db 1518 GAGTGGATGAAAGAACAGAGAAGACTTACGCCGCGCCCTTCACAAAATGACACAGTT 1577
QY 1448 TGGATGAAGAGAGAGACTTGTGCAATTCAAACATAGCAACTGAGCATTTCCAAAGCCCTG 1507
Db 1578 TGGATGAAGAGATGGGGCTCGTGCAATAGTAACACAGCAACAGAGCATATTTCCAAAGCCCTG 1637
QY 1508 CCAAGAAAGTACCAAGACTCCCTGCGCACAGCAGAGAAACCTGAAGCACTGTCAATTAAGTA 1567
Db 1638 CTAAGAAAGTGCCAAAGACTCCCTGCAACCTCGGCCGAATCTGAATCAGCTGTCAATCAGTA 1697
QY 1568 ATGGGGACATTAAGATACTCTGTGAGGCTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT 1627
Db 1698 ACGGGGAGCACTGAGAGTCTGTGGCCTTTACA-----GAGGCTCGGGGCGGGGA 1746
QY 1628 GGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGGATAATTTTAAAGATTAACATG 1687
Db 1747 TGGGGTATGGAACGCTTGAGAGAAAT-GGATGTCTTGGGACAAATTTTGCA-GATTATGTG 1804
QY 1688 TTATGTAATTTTATGTGACTGACATGAGCCTGGATGACTATCGTGTACTTTGGGAAAG 1747
Db 1805 TTGCTTAAATGTAATGTAAGTACGACAGAGCCAGAAAGCTATTTGTTTGGAAAAG 1864
QY 1748 TCT---CTTGTCTATTGCTGACATGCTTCTGTGTGTGTGGCCAAATGCCAAATGT 1804
Db 1865 TCTCTGCTGAACCTGAACCTGCTAATGCTTCTGTGTGTGTGGCCAGTGGTAATGT 1924
QY 1805 ACTCGAATGATGTT-AAAGGCTCTGTAAAACTTCATACCTCTTTGGCCATTTGTATGCAT 1863
Db 1925 ACTGCAGTGATGTTAAAGGGCTCTGTAGAACTTATACCTCTCTGGATGTTTATACGCAT 1984
QY 1864 GATGTTTGGTTTTTAAACATGGTATTAATGAATGTGTACTTCTGTCAAGAGAACAGAG 1923
Db 1985 GCAGTTTAGTTTCAGATGTGGT---ATGAACTGAGTGCTTCTGACAGCAAAAGCAGAG 2041

QY 1924 GTAATAATCTCCAATTAAAAA 1945
Db 2042 GTAATAAGTGTCCAATTTTAAA 2063
RESULT 10
BC031363
LOCUS
DEFINITION
BC031363 3278 bp mRNA linear ROD 07-OCT-2003
Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
(cDNA clone MGC:36633 IMAGE:5357048), complete cds.
ACCESSION
BC031363
VERSION
BC031363.1 GI:21618632
KEYWORDS
MGC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 3278)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Uedlin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.W., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kerteman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schererh,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 3278)
Strausberg,R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Günaratne, P.H., García, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Navevati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAK Plate: 54 Row: e Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 22122370.
FEATURES
source
1..3278
/organism="Mus musculus"
/mol_type="mRNA"

/db_xref="taxon:10090"
/clone="MGC:36633 IMAGE:5357048"
/issue_type="Eye, retina, mouse strain C57Bl\6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..3278
/gene="Hmgcs1"
/note="synonym: MGC36662"
/db_xref="locusID:208715"
/db_xref="MGI:107592"
81..1643
/codon_start=1
/product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase
1"
/protein_id="AAH31363.1"
/db_xref="GI:21618633"
/db_xref="locusID:208715"
/translation="MPSLPINAEACWPKDVGIVALEIYFPSQYVDQAELEKYGVDV
GKVTIGQARMGFCTDREDINSLCPTVQKMERHSLSYDCIGRLEVGETIIDS
SVKSNLMQLFEESGNTDIEGIDTNACYGTAVFNAVNWVESSSWDGRVALVAGDI
AIYATGNARPTGSGAVALLIGPNAPLIFDRGLRGTHMOHAYDFYKPDMLSEYPVVD
KLISQCYLSALDRCSYVRKKIRAWOMKEXKDFLNDGFMI FHSPYCKLVQKSLA
RMFLNDFLNDQNRDKNSIYSGLEAFGDKLEDTYFDRDVEKAFMKASSELFNQTKAS
LLVSNQNGMNTSSVYGSLSAVLAQYSPQLAGKRVGVFSYSGLAATLVSLKTQDA
TPGSALDKITASLCDKSRLDSTRCTVAPDVFAENMKLREDTHLANYIPQCSIDSLFE
GTWYLVRVDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATEHIPSPAKKVRPLPATSA
ESESAVISNGEH"
117..1487
misc_feature
/note="HMG CoA synt; Region:
Hydroxymethylglutaryl-coenzyme A synthase"
/db_xref="CDD:pfam01154"

ORIGIN		Query Match	59.3%;	Score 1188;	DB 10;	Length 3278;	
		Best Local Similarity	79.9%;	Pred. No. 4.2e-248;			
		Matches 1551;	Conservative	0;	Mismatches 245;	Indels 146;	Gaps 7;
OY	134	GCTCTTTCACCATGCGCTGGATCACTTCTTGAATGCAAGAGCTTGCTGGCCAAAGATG					193
Db	70	GCTCTTTCACCATGCGCTGGGTCACCTTCTTGAATGCAAGGCTTGCTGGCCAAAGATG					129
OY	194	TGGGAATTGTTGCCCTTGAGATCTATTTTCCCTTCAATATGTGATCAAGCAGATTGG					253
Db	130	TGGGAATCGTTGCCCTTGAATCTACTTCCCTTCAATATGTGATCAAGCTGAGTTGG					189
OY	254	AAAAATATGATGCTGTAGATGCTGGAAAGTATACCAATTGGCTTGGCCAGGCCAGATGG					313
Db	190	AAAAATACGATGCTGTAGATGCTGGAAAGTATACCAATCGGCTGGGCCAGGCCAGATGG					249
OY	314	GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTGCATGACTGTGTTCAGAACTTA					373
Db	250	GCTTCTGCACGGATCGTGAAGACATCAACTCTCTTGGCCGACTGTGTTCAAGAACTGA					309
OY	374	TGGAGAGAAATTAACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGACAA					433
Db	310	TGGAAGAGACATAGCCTTTCCTATGATTGCATTGGGCGGCTAGAAATTGGAACAGACAA					369
OY	434	TCATCGACAAATCAAAGTCTGTGAAGACTAATTGATGCAAGCTGTTGAAGAGCTGGGA					493
Db	370	TCATCGACAAATCGAAATCAGTGAAGTCTAATTGATGCAAGCTGTTGAGGAGTCTGGGA					429
OY	494	ATACAGATATAGAAGSAGATCGACAACTAATGCAATGCTATGAGGACACAGCTGCTCT					553
Db	430	ATACAGATATAGAAGSAGATAGATACAAACCAATGCAATGCTATGGGGACACAGCTGCACT					489
OY	554	TCAATGCTGTTAACTGGATGAGTCCAGCTCTTGGAT-----					591
Db	490	TCAATGCCGTGAAGTGGGTGCAATCCAGCTCTTGGGATGAGCAGATATGCTGTGTAATTG					549
OY	592	-----					591
Db	550	CAGAGACATTGCTATATATATGCCACAGGAATGCCAGACCTACAGGTGAGTTGAGCTG					609

OY	592	-----	-----	-----	-----	GGGCTT	CGTGG	AC	607													
Db	610	TGGCCCT	GCTAAT	TGGGCC	AAACGCT	CTCTAAT	TTTGA	CCGAGG	CGCTCC	GTGG	AC	669										
OY	608	ATATG	CAACAT	GCCCTAT	GATTTT	TACAAG	CCTGAT	ATGCTAT	CTGAAT	ATCT	TATAG	667										
Db	670	ACATG	CAGCAT	GCCCTAT	GACTTT	TACAAG	CCTGAC	TGCTCT	CCGAGT	ACCT	GTGG	729										
OY	668	ATG	AAAACT	CTCCAT	ACAGT	GTACCT	CAGT	GCATT	GACCG	CTGCT	ACTCT	727										
Db	730	ACG	AAAGCT	CTCCAT	ACAGT	GTACCT	CAGCGC	CTGAC	CGCTG	CTATT	CTGT	789										
OY	728	GCA	AAAA	GATCC	ATGCG	CAAGG	CAATG	ATAAG	ATTTT	TACCT	TGA	787										
Db	790	GCA	AAAA	GATCC	ATGCG	CAAGG	CAATG	ATAAG	ATTTT	TACCT	TGA	849										
OY	788	ATT	TGG	CTTCAT	GATCTT	CACTC	ACCAT	ATTG	TAACT	GGT	CAGA	ATCT	847									
Db	850	ATT	TGG	CTTCAT	GATCTT	CACTC	ACCAT	ATTG	TAACT	GGT	CAGA	ATCT	909									
OY	848	GGAT	GTGCT	GAATG	ACTTCC	TAA	TGAC	CAAG	ATATAG	ATAA	ATAG	TATCT	907									
Db	910	GGAT	GTGCT	GAATG	ACTTCT	TAA	TGAT	CAAA	CA	GAGAT	AAAA	CAGT	ATTT	969								
OY	908	GCC	TGAA	GCCTTT	GGA	TGTTAA	ATTA	GAA	CA	CACTA	CTTT	GATAG	AG	967								
Db	970	GAC	TGAA	GCCTTT	GGA	CGTTAA	ATTA	GAA	CA	CACTA	CTTT	GATAG	AG	1029								
OY	968	AGG	CA	TTTAT	GAA	GCTAG	CTGA	ACTCTT	CAGT	CA	AAAA	CAAA	GGCAT	CTTT	ACT	1027						
Db	1030	AGG	CA	TTTAT	GAA	GCTAG	CTGA	ACTCTT	CAGT	CA	AAAA	CAAA	GGCAT	CTTT	ACT	1089						
OY	1028	TAT	CAAT	CAAA	TGAA	TATGTA	CA	CACTT	CA	GTAT	ATGTTCC	CTTG	CA	CTG	1087							
Db	1090	TGT	TAAT	CA	GAATG	AAATAT	GTAC	ACAT	CTCTG	TAT	GTTC	CCCT	G	CT	1149							
OY	1088	TAG	CA	CAGT	ACTCA	CCCTC	AGCA	ATTAG	CAGG	AAAG	AAATTG	GAGT	TTTCT	TAT	GT	1147						
Db	1150	TGG	CA	CAGT	ACTCA	CCCTC	AGCA	ATTAG	CAGG	AAAG	AAATTG	GAGT	TTTCT	TAT	GT	1209						
OY	1148	CTG	TTG	GCTG	CCACT	CTGTA	CTCTT	TAA	AGTCA	CA	AGATG	CTAC	ACG	GGG	CTG	1207						
Db	1210	CTG	TTG	GCTG	CCACT	CTGTA	CTCTT	TAA	AGTCA	CA	AGATG	CTAC	ACG	GGG	CTG	1269						
OY	1208	CTC	TGAT	AAAT	TAAC	AGCTT	TATG	TATG	ATCA	AGCTT	GAAT	CA	AG	AACTG	1267							
Db	1270	CC	CTGAT	AAAT	TAAC	AGCTT	TATG	TATG	ATCA	AGCTT	GAAT	CA	AG	AACTG	1329							
OY	1268	G	TG	GCAC	CCAGAT	GTCTT	CGTGA	AAACAT	GAGCT	CA	AGAG	ACAC	CCCAT	CATT	GG	1327						
Db	1330	G	TG	GCAC	CCAGAT	GTCTT	CGTGA	AAACAT	GAGCT	CA	AGAG	ACAC	CCCAT	CATT	GG	1389						
OY	1328	TCA	ACTAT	ATTTCC	CA	GGGTTCA	TAG	ATTCA	CTCTT	TGA	AGAA	CGTGT	ACTTA	GT	1387							
Db	1390	CCA	ACTAT	ATTTCC	CA	GGGTTCA	TAG	ATTCA	CTCTT	TGA	AGAA	CGTGT	ACTTA	GT	1449							
OY	1388	G	G	TGAT	GA	AAAA	GCAC	AGAACT	TACG	CTCG	GGCTCC	CACT	CCAA	ATGAT	GAC	1447						
Db	1450	GAG	TGAT	GA	AAAA	GCAC	AGAACT	TACG	CTCG	GGCTCC	CACT	CCAA	ATGAT	GAC	1509							
OY	1448	TG	ATGA	AGAG	TAG	AGACTT	GTGCA	TTCA	AA	CA	TAGCA	CTGAG	CA	TAT	TCCAA	GGCC	CTG	1507				
Db	1510	TG	ATGA	AGAG	TAG	AGACTT	GTGCA	TTCA	AA	CA	TAGCA	CTGAG	CA	TAT	TCCAA	GGCC	CTG	1569				
OY	1508	CCA	AGAA	GTAC	CAAG	CTCC	TG	CA	AG	CA	AGCA	AACTG	AA	AGCA	AGCTG	TCAT	TA	AGTA	1567			
Db	1570	CTA	AGAA	GTAC	CAAG	CTCC	TG	CA	AG	CA	AGCA	AACTG	AA	AGCA	AGCTG	TCAT	TA	AGTA	1629			
OY	1568	AT	G	GGA	CA	CTTA	AGAT	CTCTG	TA	AGGT	GCA	AG	CTT	CA	GGG	TGG	GGT	GG	CA	TGG	GGT	1627
Db	1630	AC	G	GGA	CA	CTTA	AGAT	CTCTG	TA	AGGT	GCA	AG	CTT	CA	GGG	TGG	GGT	GG	CA	TGG	GGT	1678

QY	1628	GGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGATATTTTAAAGGATTCATG	1687
Db	1679	TGGGGTATGGAACGGTTGGAGGAAAT-GGATGCTTGGGACAATTTTGCA-CATTATGTG	1736
QY	1688	TTATGTAAATTTTATGTGATGTCACATGAGCCGATGATGATCATCTGTACTTTGGGAAG	1747
Db	1737	TTGCTTAAATGTATGTAACTGACACAGAGCCAGAAAGCTATGTGTTTGGAAAAG	1796
QY	1748	TCT--CTTGGCTATTTGCTGACATGCTTCCTGTGTGTGTGGCCCATGCCAATGT	1804
Db	1797	TCTCTGCTGAACCTGTAACATGCTTCTGTGTGTGTGGCCAGTGTAAATGT	1856
QY	1805	ACTCGATGATGTT-AAGGCGTCGTAAACCTTCATACCTCTTTGGCCATTTGTATGCAT	1863
Db	1857	ACTGCAGTGATGTTAAAGGCGTCGTAGAACCTTTATACCTCTCTGATGTTTATACGCAT	1916
QY	1864	GATGTTTGGTTTTTAAACATGGTATATGAATGTGTACTTCTGTGAGAGAAAGCAGAG	1923
Db	1917	GCAGTTTAGTTTTCAGATGTGT--ATGAACGTAGTGCTTCTGACAGCAAAAGGCAGAG	1973
QY	1924	GTACTAATCTCCAATTAAAAA	1945
Db	1974	GTAAGTGTGTCGAATTTTAA	1995

RESULT 11		
AX700129		
LOCUS	AX700129	3275 bp
DEFINITION	Sequence 15 from Patent EP1284298.	DNA linear PAT 03-APR-2003

KEYWORDS	
SOURCE	.
ORGANISM	Rattus norvegicus (Norway rat)
	Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE	AUTHORS	TITLE	JOURNAL
1	Brooksbank, R.A., Dixon, A.K., Lee, K. and Pinnock, R.D.	Identification and use of molecules implicated in pain	Patent: EP 1284298-A 15 19-FEB-2003;

```

FEATURES
  source
    1..3275
    /organism="Rattus norvegicus"
    /mol_type="unassigned DNA"
    /db_xref="taxon:10116"
    /note="Cytosolic 3-hydroxy 3-methylglutaryl coenzyme A
    synthase"

```

ORIGIN

Query Match	59.0%;	Score 1180.6;	DB 6;	Length 3275;
Best Local Similarity	78.3%;	Pred. No. 1.7e-246;		
Matches 1559; Conservative	0;	Mismatches 284;	Indels 148;	Gaps 6;

QY	134	GCTCTTTCACCATGCTCGATCACTTCCTTGAATGAGAAAGCTTGCGCCAAAAAGATG	193
Dp	81	GCTCTTTCACCATGCTCGGTCACTTCCTTGAATGACAGGCTTGCGCCAAAAAGATG	140
QY	194	TGGGAATGTTGCTCTGAGATCTATTTCTTCGAATATGTTGATCAAGCAGATGG	253
Dp	141	TGGGAATGTTGCTCTGAAATCTACTTCTCTCAGTATGTTGATCAAGCTGAGTTGG	200
QY	254	AAAAATATGATGTTGATGCTGGAAGTATACCATGGCTTGGGCCAGGCCAAGATGG	313
Dp	201	AAAAATACGATGTTGATGCTGGAAGTATACCATGGCTTGGGCCAGGCCAAGATGG	260
QY	314	GCTTCTGCACACAGATAGAGAAGATATTAATCTCTTTCATGACTGTGGTTCCAGAATCTTA	373
Dp	261	GCTTCTGCACAGATCGCGAAGACATCAACTCTCTTCGCTGACTGTGGTTCCAGAATCTTA	320
QY	374	TGAGAGAAATAACTTTCCTATGATTCATTGGCGGCTGGAATGGAACAGAGACAA	433

Db	321	TGAGAGAAATAGCCTTTCCCTATGACTGCATTGGGCGGCTGGAAGTCGGGACAGAGACAA	380
QY	434	TCATGCACAATCAAAGTCTGTGAAGACTAATTGTATGACGTGTTTGAAGACTGGGA	493
Db	381	TCATGCACAATCAAATCCGTGAAGTCTAATTTGATGCAGCTGTTGAGGAGTCTGGGA	440
QY	494	ATACAGATATAGAAGGAATGCAACAACCTAATGCATGCTATGAGGACACAGCTGCTGCT	553
Db	441	ATACAGATATAGAAGGAATGATATACAACTAATGCATGCTATGGGGGCACAGCCGCACTCT	500
QY	554	TCAATGCTGTTAACTGATGATGATCCAGCTCTTGGAT	591
Db	501	TCAATGCTGGAACCTGGATCGAATCCAGCTCTTGGATGACGATACGCTTTGGTAGTTG	560
QY	592	-----	591
Db	561	CAGGAGACATCGCTATATATATGCTTCAGGAAACGCCAGGCGCTACAGGTGAGTTGAGCTG	620

QY	592	-----GGGCTTCGTGGGACAC	607
Db	621	TGGCTCTGCTAATTGGCCAAATGCTCTGTAAATTTTGAACCGAGGGCTTCGTGGACAC	680
QY	608	ATATGCAACATGCCCTATGATTTTTCACAAGCCTGATATGCTAATCTGAAATATCCTATAGTAG	667
Db	681	ACATGCACAGATGCCCTACGACTTTTACAGCCCTGACATGCTCTCTGAATATACCCTGTGTTAG	740
QY	668	ATGGAANAACCTCCCATACAGTGTCTACCTCAGTCGACATTAGACCCGCTGTCTCTCTACT	727
Db	741	ATGGAANAACCTCCCATACAGTGTCTACCTCAGTCGACATTAGACCCGCTGTCTCTACT	800
QY	728	GCAAAAAAGATCCATGCCCCAGTGGCAGAAAGAGGAAATGATTAAGATTTTACCTTGAATG	787
Db	801	GCAAAAAAGATCCGGGCCCACTGGCAGAAAGAGGAAAGATTTTACCTTGAATG	860
QY	788	ATTTTGGCTTCATGATCTTTCACTCACCATATTTGTAACTGTGTTCAGAAAATCTCTAGCTC	847
Db	861	ATTTTGGCTTCATGATCTTTCACTCGCCATATCTGTAACTGTGTGACAGAAATCTCTAGCTA	920
QY	848	GGATGTTGCTGAATGACTTTCCTTAATGACCAAGATAGAGATAAAAATAGTATCTATAGTG	907
Db	921	GGATGTTCTGATGACTTTCCTTAACGATCAAAAACAGACACAAAACAGTATTTTACAGTG	980
QY	908	GCCTGGAACCTTTGGGATGTTAAATAGAAGACACCTACTTTGATAGAGATGTGAGA	967
Db	981	GGCTGAAGCCTTTGGGATGTGAATTAGAAGATACTTACTTCGACAGAGATGTGAAA	1040
QY	968	AGGCATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAAGGCATCTTTACTTG	1027
Db	1041	AGGCATTTATGAAGGCTAGTGTCTGAGCTATTTCACACAGAAAACAAAGGCATCTTTGCTTG	1100
QY	1028	TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTCGATCTGTTTC	1087
Db	1101	TATCAAAATCAAAATGGAACATGTACACATCTCTGTATACGTTCCCTTCGCTTCTGTTTC	1160
QY	1088	TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAGAAATTGAGTGTCTTCTATAGGTT	1147
Db	1161	TGGCACAGTACTCACTCACAAGTTGGCCGGGAAGAGATTGGAGTGTCTCTTACGGTT	1220
QY	1148	CTGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAAGATGCTACACCGGGTCTG	1207
Db	1221	CTGCTTGGCTGCCACTCTACTCCCTTAAAGTCACACAAAGATGCCACACAGGATCTG	1280
QY	1208	CTCTTGATTAATAAACAACAAGTTTATGTGATCTTAAATCAAGGCTTGATTTCAAGACTG	1267
Db	1281	CTCTTGACAAATAAACAACAAGTTTATGTGACCTTAAAGTCAAGGCTTGACTCAAGAACGT	1340
QY	1268	GTTGTGCAACAGATGTCTTCGCTGAAAAATGAAGCTCAGAGAGGACACCCATCATTTTG	1327
Db	1341	GTTGTGCAACAGACGTCTTTGCTGAAAAATGAAGCTCAGAGAGGACACACATCATTTAG	1400
QY	1328	TCAACTATATTTCCCAAGGTTCAATATGATTCACCTTTGAAAGAACGTGTACTTAGTTTA	1387
Db	1401	CCAACTATATTTCCCAAGTGTCAATATGATTCACCTTTGAAAGAACATGTGTATCTAGTTCA	1460

QY	1388	GGGTGATGAAAAAGCACAAGAAACTTACGCTCGGCGCTCCCACTCCAAATGATGACACTT	1447
Db	1461	GAGTGGATGAAAAAGCACAAGAACTTACGCGCCGCGCTCCCAAAATGACCACAGTT	1520
QY	1448	TGGATGAAGAGTAGGACTGTGTGCAATTCAAACATAGCAACTGAGCATATTCCAAAGCCCTG	1507
Db	1521	TGGATGAAGAGTAGGACTGTGTGCAATTCAAACACAGCTACAGAGCATATTCCAAAGCCCG	1580
QY	1508	CCAAGAAAGTACCAAGACTCCCTGCCCACAGCAGCAGAACTGAAGCAGCTGTCAATTA	1567
Db	1581	CTAAGAAAGTGCCCAAGACTTCTTCGCACTCCGGCCGAACCTGAGTCGGCTGTCACTA	1640
QY	1568	ATGGGGAACATTAAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGG	1627
Db	1641	ACGGGAGCAGCTGAGAGCCGTGGCCTTCACAGAGGCTCGGGCT-----GGA	1687
QY	1628	GGGGGTATGGGAACAAGTTGGAAGAAATGGGATATCTGGGGATTAATTTTAAAGATTACA	1687
Db	1688	TGGGTATGGGAACGTTAGAGGAAT-GGATGCTTGGGACAATTTTACAGATTACGTGT	1746
QY	1688	TTATGTAATTTTATGTGACTGACATGAGAGCCGTGAAGACTATCGTACTTGGGAAAG	1747
Db	1747	TGCTTAAATGTATGTAACTGACACAGAGCCCAAGAAACTGTGTGTTTGGAGAG	1806
QY	1748	TCTCTTGTCTTATTTGCTGACATGCTTCCCTGTGTGTGTGGCCCAATGCCAATGTACT	1807
Db	1807	TCTCGCTG---AGCTCCTAACACACACTTCTGTGTGGCTGGCCAATGTGAATGTACT	1862
QY	1808	CGAATGATGTTAAGGCGCTGTAAAACTTCATACCTCTTTGGCCATTTGTATGATGATG	1867
Db	1863	GCGATGTTGTTAAGGCGCTGCAAGACGTACATCCTGCTGCATGTTTACACGCATGCG	1922
QY	1868	TTTGTTTTTAAACATGCTATATGAATTGTACTTCTGTCAAGAAAGCAGAGTAC	1927
Db	1923	GTTAGGCTTCAAACTCGGT---CTGAAGTGAAGTGTCTGACTGCA-AAGCAGAGGTAC	1978
QY	1928	TAATCTCCAATTAAAAATTTTAAACATGTAAAAAATAAAAAAAAAAAAAAAAAAAAA	1987
Db	1979	TGCTGTCCAGTTAAAAAATTGTTTTTTTTTTTAAATGTGTAAAGATTTTTATACTTA	2038
QY	1988	AAAAAAAAAAAAA 1998	
Db	2039	AATAAAAAAA 2049	
RESULT 12			
AX827303			
LOCUS	AX827303	3275 bp	DNA linear PAT 12-DEC-2003
DEFINITION	Sequence 37 from Patent EP1344834.		
ACCESSION	AX827303		
VERSION	AX827303.1 GI:39837392		
KEYWORDS			
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1		
AUTHORS	Boess, F., Suter-Dick, L. and Wolf, D.		
TITLE	Methods for the toxicity prediction of a compound		
JOURNAL	Patent: EP 1344834-A 37 17-SEP-2003;		
	F. HOFFMANN-LA ROCHE AG (CH)		
FEATURES	Location/Qualifiers		
source	1..3275		
	/organism="Rattus norvegicus"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:10116"		
ORIGIN			
Query Match	59.0%;	Score 1180.6;	DB 6; Length 3275;
Best Local Similarity	78.3%;	Pred. No. 1.7e-246;	
Matches 1559; Conservative	0;	Mismatches 284;	Indels 148; Gaps 6;

QY	134	GCTCTTTCACCATGCCCTGGATCACTTCCCTTGAATGCAGAAAGCTTGCTGGCCAAAAGATG	193
Db	81	GCTCTTTCACCATGCCCTGGGTCACTTCCCTTGAATGCAGAGGCTTGCTGGCCAAAAGATG	140
QY	194	TGGGAATGTGTCCTTGAGATCTATTTTCTCTCAATATGTTGATCAAGAGATTGG	253
Db	141	TGGGAATCGTTGCCCTTGAAATCTACTTTCCTTCTCAGTATGTTGATCAAGCTGAGTTGG	200
QY	254	AAAAATATGATGGTGTAGATGCTGGAAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG	313
Db	201	AAAAATACGATGGTGTAGATGCTGGAAAGTATACCATTTGGCCCTGGGCCAGGATGG	260
QY	314	GCTTCTGCACAGATAGAGAATATTAACCTCTTTGCATGACTGTGGTTCAGAACTTCA	373
Db	261	GCTTCTGCACGGATCGCGAAGACATCAACTCTCTGCTGACTGTGGTTCAGAACTGA	320
QY	374	TGGAGAGAAATAAAGCTTTCCTATGATTCATTGGCCGGCTGGAAAGTTGGAACAGAGACAA	433
Db	321	TGGAGAGAAATAGCCTTTCCTATGACTGCATTGGCGGCTGGAAGTCGGGACAGAGACAA	380
QY	434	TCATCGACAAATCAAGTCTGTGAAGACTAAATTGATGCAGCTGTTTGAAGAGTCTGGGA	493
Db	381	TCATCGACAAATCAAAATCCGTGAAGTCTAAATTGATGCAGCTGTTTGAAGAGTCTGGGA	440
QY	494	ATACAGATATAGAAAGAAATCGACACAACTAATGCATGCTATGGAAGCACAGCTGCTCT	553
Db	441	ATACAGATATAGAAAGAAATAGATACAACTAATGCATGCTATGGAAGCACAGCCGAGTCT	500
QY	554	TCAATGCTGTAACTGGATTGAGTCCAGCTCTTGGGAT-----	591
Db	501	TCAATGCTGTGAAGTGAATCGAATCCAGCTCTTGGGATGGACGATACGCTTGGTAGTTG	560
QY	592	-----	591
Db	561	CAGAGACATCGCTATATATGCTCAGGAAAGCCAGCCTACAGGTGAGTTGAGCTG	620
QY	592	-----GGGCTTGTGGGACAC	607
Db	621	TGGCTCTGCTAATTGGGCCCAATGCTCCTGTAAATTTTGACCGAGGGCTTCGTGGGACAC	680
QY	608	ATATGCAACATGCTTATGATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGTAG	667
Db	681	ACATGCAAGATGCTTACGACTTTTACAAAGCCTGACATGCTCTCTGAATATCCTGTGTAG	740
QY	668	ATGGAAGAACTCTCCATACAGTGTCTACCTCAGTGCATTTAGACCCTGCTACTCTGTACT	727
Db	741	ATGGAAGAACTCTCCATACAGTGTCTACCTCAGCGCATTTGAAACGCTGCTATTCTGTCTACC	800
QY	728	GCAAAAAAGATCCATGCCCACTGGCAGAAAGAGGGAATGATAAAGATTTTACCTTGAATG	787
Db	801	GCAAAAAAGATCCGGGCCCACTGGCAGAAAGAGGGAATGAAGATTTTACCTTGAATG	860
QY	788	ATTTGGCTTCATGATCTTTCACCTACCACTAATTTGTAACCTGGTTCAGAAATCTTACGCTC	847
Db	861	ATTTGGCTTCATGATCTTTCACCTCGCCATATCTGTAACTGGTGCAAGAAATCTTACGCTA	920
QY	848	GGATGTTGCTGAATGACTTCTTAATGACAGAAATAGAGATAAAATAGTATCTATAGTG	907
Db	921	GGATGTTTCCGAAATGACTTCTTAAACGATCAAAACAGAGACAAAAACAGTATTTACAGTG	980
QY	908	GCCTGGAAGCCTTGGGGATGTTAAATTAGAGACACTTCTTGTATAGAGATGTGAGA	967
Db	981	GGCTGGAAGCCTTGGGGATGTGAATATTAGAGATACTTACCTCGACAGAGATGTGAGAA	1040
QY	968	AGGCATTTATGAAGGCTAGCTCTGAAGCTTCACTCAGTCAAGAAACAAAGGCATCTTACTTG	1027
Db	1041	AGGCATTTATGAAGGCTAGTGCTGAGCTATTCAACAGAAAAACAAAGGCATCTTGTCTTG	1100
QY	1028	TATCAAAATCAAAATGGAATATATGTACACATCTTCAGTATATGTTCCCTTGCACTCTGTTCC	1087
Db	1101	TATCAAAATCAAAATGGAACATGTACACATCTCTGTATACGGTTCCTTGTCTCTGTTCC	1160

QY	1088	TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAAATTGGAGTGTCTTATAGTT	1147
Db	1161	TGGCACAGTACTCACTCAACAGTTGGCCGGGAAGAGATTGGAGTGTCTTACGGTT	1220
QY	1148	CTGGTTGGCTGGCCACTCTGTACTCTTTAAAGTCAACACAGATGCTACACCGGGGCTG	1207
Db	1221	CTGGCTTGGCTGGCCACTCTACTCTCTTAAAGTCAACACAGATGCCACCGAGATCTG	1280
QY	1208	CTCTTGATAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG	1267
Db	1281	CTCTTGACAAATAACAGCAAGTTTATGTGACCTTAAAGTCAAGGCTTGACTCAGAACGT	1340
QY	1268	GTGTGGCACCAAGATGTCTTCGCTGAAAAACATGAGCTCAGAGAGACACCCATCATTTGG	1327
Db	1341	GTGTGGCACCAAGCGTCTTGTGAAAAACATGAGCTCAGAGAGACACACATCATTTAG	1400
QY	1328	TCAACTATATTCCTCCAGGGTTCATAGATTCACTCTTTGAAGGAACGTGTACTTAAGTTA	1387
Db	1401	CCAACATATTCCTCCAGTGTTCATAGATTCACTCTTCGAAGGAACATGTATCTAGTCA	1460
QY	1388	GGGTGATGAAAAACACAGAAGACTTACGCTCGCGCTCCCACTCCAATGATGACACTT	1447
Db	1461	GAGTGATGAAAAACACAGAAGACTTACGCCCCGCGCTCCCTCCACAAATGACCAAGTT	1520
QY	1448	TGGATGAAGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTCCAAAGCCCTG	1507
Db	1521	TGGATGAAGAGTAGGACTTGTGCAATTCAAACACAGCTACAGAGCATATTCCAAAGCCCGG	1580
QY	1508	CCAAAGAAAGTACCAAGACTCCCTGCCACACAGCAGACGAACCTGAAGCAGCTGTCAATTA	1567
Db	1581	CTAAGAAAGTCCCAAGACTTCTCGCGACCTCGGGCGAACCCTGAGTGGCTGTCACTCA	1640
QY	1568	ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT	1627
Db	1641	ACGGGGAGACACTGAGAGCCGTGGCTTCACAGAGGCTCGGGCT-----GGA	1687
QY	1628	GGGGGTATGGGAACAGTTGAGGAATGGATATCTGGGGATAATTTTAAAGATTACATG	1687
Db	1688	TGGGGTATGGGAACGTTAGAGGAAT-GGATGTCTTGGGACAATTTTACAGATTACGTGT	1746
QY	1688	TTATGTAAATTTTATGTGACTGACATGAGGCTTGATGACTATCGTGTACTTGGGAAAG	1747
Db	1747	TGCTTAAAAATGTAAATGTAACTGACACAGAGCCCAAAAACTGTGTGTTTGGAGAAG	1806
QY	1748	TCTCTTGTCTTATTTGCTGACATGCTTCTGTGTGTGCTTGCCCAATGCCAAATGTACT	1807
Db	1807	TCTCGCTG---AGCTCCTAACACACTTCTGTGTGTGGCTGGCCAATGGTGAATGTACT	1862
QY	1808	GCAATGATGTTAAGGGCTCTGTAAAACTTCATACCTCTTTGGCCAATTGTATGCATGATG	1867
Db	1863	GCGATGCTGTTAAGGGCTCTGCAAGAACGTCATACCTCGCTGCATGTTTACACCGCATGCCG	1922
QY	1868	TTTGGTTTTTAAACATGGTATATGAATTTGTACTTCTGTGCAGAAAGAAAGCAGAGGTAC	1927
Db	1923	GTTAGGCTTCAAACCTCGT---CTGAAGTGAAGTCTTCTGACTGCA-AAAGCAGAGGTAC	1978
QY	1928	TAATCTCCAATTAAAAAATTTTAAACATGTMAAAAAAATAAAAAAATAAAAAA	1987
Db	1979	TGCTGTCCAGTTTAAAAAATGTTTTTTTTTTTTTAAATGTATAAGAAATTTTAACTTA	2038
QY	1988	AAAAAAAAAAAAA	1998
Db	2039	AATAAAAAAA	2049

RESULT 13
RNCHMCA
LOCUS RNCHMCA 3275 bp mRNA 1linear ROD 18-JUN-2003
DEFINITION Rattus norvegicus mRNA for cytosolic 3-hydroxy 3-methylglutaryl
coenzyme A synthase (EC 4.1.3.5).
ACCESSION X52625
VERSION X52625.1 GI:55946
KEYWORDS cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase.

SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 Ayte,J., Gil-Gomez,G. and Hegardt,F.G. Nucleotide sequence of a rat liver cDNA encoding the cytosolic 3-hydroxy-3-methylglutaryl coenzyme A synthase Nucleic Acids Res. 18 (12), 3642 (1990)
JOURNAL	MEDLINE 90301491 1972979 (b 1-38,40,41-62,64-3275)
REMARK	2 Hegardt,F.G.
AUTHORS	Direct Submission Submitted (17-APR-1990) Hegardt F.G., University of Barcelona, Unit of Biochemistry, School of Pharmacy, Plaça Pius XII, s/n. 08028 Barcelona, Spain
TITLE	Journal
REMARK	3 (bases 1 to 3275) Hegardt,F.G.
AUTHORS	Direct Submission Submitted (30-JUL-1990)
TITLE	Journal
REMARK	1. Location/Qualifiers 1. .3275 /organism="Rattus norvegicus" /mol_type="mRNA" /strain="Sprague Dawley" /db_xref="taxon:10116" /clone="lambda-ccs1" /issue_type="liver" 31. .70 /note="revised by [3]" /citation=[3] 92. .1654 /EC_number="4.1.3.5" /codon_start=1 /product="cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase"
old_sequence	31. .70 /note="revised by [3]" /citation=[3] 92. .1654 /EC_number="4.1.3.5" /codon_start=1 /product="cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase"
CDS	1. Location/Qualifiers 1. .3275 /organism="Rattus norvegicus" /mol_type="mRNA" /strain="Sprague Dawley" /db_xref="taxon:10116" /clone="lambda-ccs1" /issue_type="liver" 31. .70 /note="revised by [3]" /citation=[3] 92. .1654 /EC_number="4.1.3.5" /codon_start=1 /product="cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase"
ORIGIN	Query Match 59.0%; Score 1180.6; DB 10; Length 3275; Best Local Similarity 78.3%; Pred. No. 1.7e-246; Matches 1559; Conservative 0; Mismatches 284; Indels 148; Gaps 6; 134 GCTCTTACCAATGCTGGATGACTCTCTTGAATGAGAAAGCTTGCCCAAAAGATG 193 81 GCTCTTACCAATGCTGGATGACTCTCTTGAATGAGAAAGCTTGCCCAAAAGATG 140 194 TGGGAATGTTGGCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGAGTTGG 253 141 TGGGAATGTTGGCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGAGTTGG 200 254 AAAAATATGATGTTGATGCTGGAAGATATACCATTTGGCTTGGGCCAAGCAAGATGG 313 201 AAAAATATGATGTTGATGCTGGAAGATATACCATTTGGCTTGGGCCAAGCAAGATGG 260 314 GCTTGTGACAGATGAGAAAGATATTAATCTCTTTCATGACTGTGGTTCAGAAATCTTA 373

Db	261	GCTTCTGCACGGATCGCGAAGACATCAACTCTCTCTGCCCTGACTGTGGTTCAGAAACTGA	320
Qy	374	TGAGAGAAATAACCTTTTCCATATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAA	433
Db	321	TGGAGAGAAATAGCCTTTCTTATGACTGCATTGGGCGGCTGGAAGTCGGGACAGAGACAA	380
Qy	434	TCATCGACAAATCAAAGTCTGTGAAGACTAATTGTATGCAGCTGTTGAAGAGTCTGGGA	493
Db	381	TCATCGACAAATCAAATCCGTGAAGTCTAATTGTATGCAGCTGTTGAGGAGTCTGGGA	440
Qy	494	ATACAGATATAGAAGGAATGCACACAACCTAATGCATGCTATGGAGCACAGCTGCTGCT	553
Db	441	ATACAGATATAGAAGGAATGATACACTAATGCATGCTATGGGGCACAGCCGCACTCT	500
Qy	554	TCAATGCTGTTAACTGGATTGAGTCCAGCTCTGGGAT-----	591
Db	501	TCAATGCTGTGAACCTGAGTCGAATCCAGCTCTTGGAGTGCAGATAAGCTTTGTAGTTG	560
Qy	592	-----	591
Db	561	CAGGAGACATCGCTATATATAGCTCAGGAAACGCCAGCCTACAGGTGAGTTGAGCTG	620
Qy	592	-----GGGCTTCGTGGACAC	607
Db	621	TGGCTCTGCTAATTTGGGCCAAATGCTCCTGTATTTTGAACCGAGGGCTTCGTGGACAC	680
Qy	608	ATATGCACATGCTTATGATTTTTCACAGCCTGATATGCTATCTGAATATCCTATAGTAG	667
Db	681	ACATGCAGCATGCTTACGACTTTTTCACAGCCTGACATGCTCTGTAATACCTGTGTAG	740
Qy	668	ATGAAAACTCTCCATATACGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT	727
Db	741	ATGAAAACTCTCCATATACGTGCTACCTCAGCAGCATGACCGCTGCTATTCTGTCTACC	800
Qy	728	GCAAAAAGATCCATGCCCAAGTGCAGAAAAGAGGAAATGATTAAGATTTTACCCTGAATG	787
Db	801	GCAAAAAGATCCGGGCCAAGTGCAGAAAAGAGGAAAGATTAAGATTTTACCCTGAATG	860
Qy	788	ATTTTGGCTTCATGATCTTTCACCTCACCATATTTGTAAACTGTGTCAGAAATCTTAGCTC	847
Db	861	ATTTTGGCTTCATGATCTTTCACCTCACCATATTTGTAAACTGTGTCAGAAATCTTAGCTA	920
Qy	848	GGATGTTGCTGAATGACTTCTTAATGACCAGAAATAGATAAATAAGTATCTATAGTG	907
Db	921	GGATGTTGCTGAATGACTTCTTAACGATCAAAACAGAGCAAAAACAGTATTTACAGTG	980
Qy	908	GCCGTGAAGCCTTTGGGAGTGTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA	967
Db	981	GGCTGGAAGCCTTTGGGAGTGTGAATTAAGAGATACTTACCTCGACAGAGATGTGAAA	1040
Qy	968	AGGCATTTATGAAGGCTAGCTCTGAACCTCTCAGTCAGAAAACAAGGCATCTTACTTG	1027
Db	1041	AGGCATTTATGAAGGCTAGCTGTGAGCTATTCAACCAGAAAAACAAGGCATCTTGTCTTG	1100
Qy	1028	TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATAGTTCCTTGCACTGTTC	1087
Db	1101	TATCAAAATCAAAATGGAACATGTACACATCCTCTGTATACGTTCCCTTGCTTCTGTTTC	1160
Qy	1088	TAGCACAGTACTCACTCTCAGCAATTAAGCAGGGAAGAGAAATGAGTGTCTTATGTT	1147
Db	1161	TGGCACAGTACTCACTCTCAGCAAGTTGGCCGGGAAGAGATGGAAGTCTCTTACGTT	1220
Qy	1148	CTGCTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAGATGCTACACCGGGGCTTG	1207
Db	1221	CTGCTTGGCTGCCACTCTACTCTCCCTTAAAGTCACACAGATGCCACACAGAGATCTG	1280
Qy	1208	CTCTTGATAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG	1267
Db	1281	CTCTTGACAAAATTAACAGCAAGTTTATGTGACCTTAAAGTCAAGGCTTGACTCAAGAACGT	1340
Qy	1268	GTGTGGCACCAGATGTCTTGCCTGAAAAACATGAAGCTCAGAGAGACACCCATCATTTGG	1327

Db	1341	GTGTGGCACCAGACGTCTTTGGCTGAAAAACATGAAGCTCAGAGAGACACACATCACTTAG	1400
QY	1328	TCAACTATATTTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA	1387
Db	1401	CCAACTATATTTCCCAAGTGTTCATAGATTCACTCTTCGAAGGAACATGGTATCTAGTCA	1460
QY	1388	GGGTGATGAAAAACAAGAGAACCTTACGCTCGGCGTCCCACTCCAAATGATGACACTT	1447
Db	1461	GAGTGATGAAGAACAAGAGAACCTTACGCCCGCGCTCCCTCCAAATGACACAGTT	1520
QY	1448	TGGATGAAGAGTAGGACTTGTGCATTCAACATAGCAACTGAGCATATTCGAAGCCCTG	1507
Db	1521	TGGATGAAGAGTAGGGAAGTGTGCATTCAACACAGCTACAGAGCATATTCGAAGCCCGG	1580
QY	1508	CCAAGAAAGTACCAAGACTCCCTGCCACAGACAGAAACCTGAAGCAGCTGTATTAGTA	1567
Db	1581	CTAAGAAAGTCCCAAGACTTCTCGCAGCTCGGGCGAACTTGAGTCGGCTGTCACTAGTA	1640
QY	1568	ATGGGGAACATTAAAGATACTCTGTGAGGTGCAAGACTTCAGGGGTGGGGTGGGCATGGGT	1627
Db	1641	ACGGGGAGCACTGAGAGCCGCTGCTTACAGAGGGCTCGGGGCT-----GGA	1687
QY	1628	GGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGATATTTAAAGATTACATG	1687
Db	1688	TGGGGTATGGGAACGTTAGAGGAAT-GGATGTCTTGGACAATTTTACAGATTACGTGT	1746
QY	1688	TTATGTAATTTTATGTGACTGACATGAGCCTGGATGACTATCGTGTACTTGGGAAG	1747
Db	1747	TGCTTAAAAATGTAATGTAACTGACACAGAGCCAGAAAACTGTTGTGTTTGGAGAAG	1806
QY	1748	TCTCTTTGCTTATTTGCTGACATGCTTCTGTTGTGCTGAGCAATGCCAATGTACT	1807
Db	1807	TCTCGCTG---AGCTCCTAACACACCTTCTGCTGTGGGCTGGCCAAATGGTGAATGTACT	1862
QY	1808	CGAATGATGTTAAGGGCTCTGTAAAACTTCATACCCTCTTGGCCATTGTTATGCATGATG	1867
Db	1863	GCGATGGTGTAAAGGGCTCTGCAGAACGTCATACCTCGCTGCATGTTTACACGCATGCGG	1922
QY	1868	TTTGGTTTTTAAACATGCTATATGAATTGTACTTCTGTGAGAAAGACAGAGGTAC	1927
Db	1923	GTTAGGCTTCAAACTCGGT--CTGAACGTAGTGCTTCTGACTGCA-AAAGCAGAGGTAC	1978
QY	1928	TAATCTCCAATTAAAAAATTTTAAACATGTAAAAAATAAAAAAATAAAAAAATAAAAAA	1987
Db	1979	TGCTGTCCAGTTTAAAAAATTTGTTTTTTTTTTTTTAATGTGTAGAAATTTTATACTTA	2038
QY	1988	AAAAAAAAAAAAA 1998	
Db	2039	AATAAAAAAA 2049	

RESULT 14

AL356361

LOCUS

DEFINITION AL356361 127832 bp DNA linear PRI 20-JUN-2002

Human DNA sequence from clone RP11-35L17 on chromosome 1, complete sequence.

ACCESSION AL356361

VERSION AL356361.21 GI:21537425

KEYWORDS

SOURCE HTG.

ORGANISM Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 127832)

Lloyd, C.

Direct Submission

Submitted (31-MAY-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

On Jun 21, 2002 this sequence version replaced gi:18693024.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RP11-35L17 is from the library RPc1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6.

FEATURES
 source
 1..127832
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-35L17"
 /clone_lib="RPc1-11.1"

ORIGIN

Query Match	55.1%;	Score 1103.2;	DB 9;	Length 127832;
Best Local Similarity	79.4%;	Pred. No. 1.2e-229;		
Matches 1465;	Conservative 0;	Mismatches 198;	Indels 183;	Gaps 6;
Db	245	CAGAGTTGAAAAATATGATGCTGTAGATGCTGGAAAAGTATACCATTTGGCTTGGGCCAGG	304	
Db	80799	CAAGTTGAAAAAGTATGTTATATAGATGCTGGAAAAGTATACCATTTGGCTTGGGCCAGG	80858	
Qy	305	CCAAGATGGGCTTCTGCACAGATAGAGAGATTAATTAATCTCTTTGGCATGACTGTGTTTC	364	
Db	80859	CCAAGATGGGCTCTGCACACATGGAGAGAGATCAACTCTCTTTGCATGCTGTGTTTC	80918	
Qy	365	AGAATCTTATGAGAGAAATACCTTTCCTATGATTTGCATTGGCGGCTGGAAGTTGGAA	424	
Db	80919	AGAATCTTATGAGAGAAAAATCTTTCCTACGATTGCATCTGGCCGACTGGAAGTTGGAA	80978	
Qy	425	CAGAGCAATCATCGACAATCAAGCTGTGAGAGACTAATTTGATGACGCTGTTGAAG	484	
Db	80979	CAGAGCAATCAAGACAATCAAGTCAAGTGAAGACTAATTTGATGACGCTGTTGAAG	81038	
Qy	485	AGTCTGGAATCAGATATAGAGGAATCGACACAATAATGATGATGAGGCAAG	544	
Db	81039	AGTCTGGGATATAGATATAGAGGAACCGACACAATCATGATGCTATGAGCCCCAG	81098	
Qy	545	CTGCTGTCTTCAATGCTGTTAACTGATTTGAGTCAAGCTCTTGGGAT-----	591	
Db	81099	CTGCTGTCTTCAAGTGTGTTAACTGATTTGAGTCAAGCTCTTGGGATGACAGTTTGGCC	81158	
Qy	592	-----	591	
Db	81159	TGGAGTTTGTAGAGATATTGCTGTATATCCACAGAGAAAGTTAGACTGACAGTAAAG	81218	
Qy	592	-----GGGCTTC	598	
Db	81219	TCAGAGCAGTAGCTCTGCTAATTGAGCCAAATGCACTTAATTTTGAACGAGGGCTTC	81278	
Qy	599	GTGGGACACATATGCAACATGCTCTATGATTTTACAAAGCCTGATATGCTATCTGAATATC	658	
Db	81279	GTGGGACACATATGCAACGCTGCTATGGCTTTTATGAGCCTGATATGCTCTCTGAATGTC	81338	

Qy	659	CTATAGTAGATGAAAACTCTCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGTACT	718	
Db	81339	CTATAGTAGATGAAAACTCTCCACACAGTGTCTCTCAGTGCATTAGATCACTGTCTATT	81398	
Qy	719	CTGTCTACTGCAAAAAGATTCATGCCAGTGGCAGAAAAGAGGAAATGATAAGATTTTA	778	
Db	81399	CTGTGACTGCAAAACGATCCGTGCCAGTGGCAGAAAAGCGAAATGATAAGATTTTA	81458	
Qy	779	CCTTGATGATTTTGGCTTCATGATCTTTCACTCACCATATTTGTAACCTGGTCAGAAT	838	
Db	81459	CCTTGATATATTTTGGCTTCTTGATCCTTCATCTCACCATATTTGTAACCTGGTCAAT	81518	
Qy	839	CTCTAGCTGGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAGATAAAATAGTA	898	
Db	81519	CTCTAGCTGGATGTTGCGGAATGACTTCTTAATGATCAGAACAGAGATAAAATAGTG	81578	
Qy	899	TCTATAGTGGCCTGGAAGCCTTTGGGATGTTAAATTAGAAAGCACTTCTTGATAGAG	958	
Db	81579	TTTGAATGCTCTGGAAGCCTTTGGGAATGTTAAATTAGAAAGCATCTATTTTGATACAG	81638	
Qy	959	ATGTGAGAGAGCATTATGAGAGGCTAGCTCTGAACCTCTTCAGTCAGAAAACAAAGCAT	1018	
Db	81639	ATGTGAGAGAGCATTATGAGAGATAGATGTTCCGAACCTCCCATCAGAAAACAAAGCAT	81698	
Qy	1019	CTTTACTTGTATCAAAATCAAAATGGAATATGTAACATCTTCATATATGTTCCCTTG	1078	
Db	81699	CTTTGC-TGATTCAGTGAAAAATGGAATATGTAACATCTTCATATATGTTCCCTTG	81756	
Qy	1079	CATCTGTTCTACACAGTACTCTCAGCAATTAAGAGGAGAGAGAAATGGAGTGTTTT	1138	
Db	81757	CAACTGTCCAGCGCAGTACTCACTCAGCAATGATCAGGAGAGAG-ACTGAGTGTCT	81815	
Qy	1139	CTTATGTTCTGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAGATGCTACAC	1198	
Db	81816	CGTACGGTCTGTTTCAAGCCGCCACCTGTACTCTCTTAAAGTCCACAGATGCCACAC	81875	
Qy	1199	CGGGTCTGCTTGTATAAATAACAGCAAGTTATGTGATCTTTAAATCAAGCCTGATT	1258	
Db	81876	CATGTTCTGCTGTGATAAATAACAGCAAGCTTATGGATCTTTAAATCAAGGTTAGCT	81935	
Qy	1259	CAAGACTGGTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGCACCC	1318	
Db	81936	CAAGACTGTGTGTGTACAGATGTCTTGTGAAAAACATGAAGCTCAGAGAGCACCC	81995	
Qy	1319	ATCATTTGTCAACTATATTTCCCGAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTT	1378	
Db	81996	ATCATTTGTCAACTCTGTTCCCGGGGCCAAGAGATTCACTCTTTGAAGGAACATGTT	82055	
Qy	1379	ACTTAGTTAGGGTGGATGAAAAGCAGACAGAACTTAACGCTGGCGCTCCACTCCAAATG	1438	
Db	82056	ACTTAGCTATGTGGTGGTGAAGAGCAGAGAAAGTTTCACTCCGTTCTCTCCAAATA	82115	
Qy	1439	ATGACACTTTTGATGAAGGAGTAGACTTGTGCAATTTCAACATAGCACTGAGCATATTC	1498	
Db	82116	ATGACACTTTGGATGAAGGAGTAGACTTGTGCAATTTCAACACAGTAACTGATCATATTC	82175	
Qy	1499	CAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGAGAACTGAAGCAGCTG	1558	
Db	82176	CAAGCCCTGCTAAGACAGTGCAGAACTCCCTGCCACAGCAGAGAAACGGAGCAGCTG	82235	
Qy	1559	TCAATTAGTAATGGGGAACATTAAGATACTCTGTAGAGGTGCAAGACTTCAAGGTGGGTG	1618	
Db	82236	TCAATTAGTAATGGGGAACATTAAGATATTCTGTAGAGGTGCAAGCTTGTGTGAGGTG	82295	
Qy	1619	GCATGGGGTGGGGTATGGAAAGATTGAGGAATGGGATATCTGGGATAATTTTAAAG	1678	
Db	82296	G-----GGATGG	82302	
Qy	1679	GATTACATGTTATGTAAATTTTATGTGACTGACATGAGCCTGGATGACTATCGTTAC	1738	
Db	82303	GATTACATGTTGCTCAGATTTTATGTGAGTGAATGAGAGCCTGGAATGACCATCGTTAC	82362	
Qy	1739	TTGGGAAAAGTCTCT-TTGCTTAATTTGCTGACATGCTTCTGTTGTGCTGGCCAATGC	1797	

|||||
Db 82363 TTGGGAAAGTCTCTTTCCATATTGCTGACATGTTCCCACTGTGTGGCCAATGC 82422
QY 1798 CAAATGTACTCGAATGATGTTAAGGGCTCTGTAAACTTCATACCTCTTTGGCCATTGT 1857
Db 82423 CAAGTCCGCTTGAGTGATGTTAAGGGCTCTATAAACTTCAGACCTCTAGCCATTGCT 82482
QY 1858 ATGCATGATGTTTGGTTTAAACATGCTATATGATGAATGCTGCTACTTCTGTGAGAGAAA 1917
Db 82483 ATACATGAAGTTAGTTTCAACCATGTGATATGAATTCCTGCTTCTGTGAGAGGAAA 82542
QY 1918 GCAGAGTACTAATCTCCAATTAAAAATTTTAAACATGTAAAA 1963
Db 82543 GCAGAGTACTAATCTTCATTAAATTTTAAACAGATAAGAA 82588

RESULT 15
AC021792/c
LOCUS AC021792 170834 bp DNA linear HTG 04-MAY-2001
DEFINITION Homo sapiens clone Rp11-661L5, WORKING DRAFT SEQUENCE, 29 unordered
pieces.
AC021792 AC021792.3 GI:8072559
VERSION AC021792.3 GI:8072559
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Birren,B., Linton,L., Nussbaum,C. and Lander,E.
TITLE 1 (bases 1 to 170834)
JOURNAL Homo sapiens, clone Rp11-661L5
REFERENCE
AUTHORS 2 (bases 1 to 170834)
JOURNAL Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
ChoepeI,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pieree,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7463245.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4319
Center clone name: 661_L_5
----- Summary Statistics
Sequencing vector: M13; M77815; 99% of reads
Sequencing vector: Plasmid; n/a; % -0.fff of reads
0.524475524475524Chemistry: Dye-terminator Big Dye, 100% of
reads
Assembly program: Phrap; version 0.960731
Consensus quality: 149820 bases at least Q40
Consensus quality: 160100 bases at least Q30

Consensus quality: 164767 bases at least Q20
Insert size: 174000; agarose-fp
Insert size: 168034; sum-of-contigs
Quality coverage: 3.2 in Q20 bases; agarose-fp
Quality co.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 1608 1607: contig of 1607 bp in length
* 1708 1707: gap of 100 bp
* 2937 2937: contig of 1230 bp in length
* 2938 3037: gap of 100 bp
* 3038 4551: contig of 1514 bp in length
* 4552 4651: gap of 100 bp
* 4652 5804: contig of 1153 bp in length
* 5805 5904: gap of 100 bp
* 5905 7016: contig of 1112 bp in length
* 7017 7116: gap of 100 bp
* 7117 8323: contig of 1207 bp in length
* 8324 8423: gap of 100 bp
* 8424 10330: contig of 1907 bp in length
* 10331 10430: gap of 100 bp
* 10431 12576: contig of 2146 bp in length
* 12577 12676: gap of 100 bp
* 12677 15039: contig of 2363 bp in length
* 15040 15139: gap of 100 bp
* 15140 17696: contig of 2557 bp in length
* 17697 17797 17796: gap of 100 bp
* 17797 20417: contig of 2621 bp in length
* 20418 20517: gap of 100 bp
* 20518 23021: contig of 2504 bp in length
* 23022 23121: gap of 100 bp
* 23122 25657: contig of 2536 bp in length
* 25658 25757: gap of 100 bp
* 25758 29388: contig of 3631 bp in length
* 29389 29488: gap of 100 bp
* 29489 34207: contig of 4719 bp in length
* 34208 34307: gap of 100 bp
* 34308 38295: contig of 3988 bp in length
* 38296 38396: gap of 100 bp
* 38396 42889: contig of 4494 bp in length
* 42890 42989: gap of 100 bp
* 42990 47895: contig of 4906 bp in length
* 47896 47996: gap of 100 bp
* 47996 54590: contig of 6595 bp in length
* 54591 54690: gap of 100 bp
* 54691 59447: contig of 4757 bp in length
* 59448 59547: gap of 100 bp
* 59548 65836: contig of 6289 bp in length
* 65837 65936: gap of 100 bp
* 65937 72591: contig of 6655 bp in length
* 72592 72691: gap of 100 bp
* 72692 83266: contig of 10575 bp in length
* 83267 83366: gap of 100 bp
* 83367 94803: contig of 11437 bp in length
* 94804 94903: gap of 100 bp
* 94904 110268: contig of 15365 bp in length
* 110269 110368: gap of 100 bp
* 110369 124647: contig of 14279 bp in length
* 124648 124747: gap of 100 bp
* 124748 140934: contig of 16187 bp in length
* 140935 141034: gap of 100 bp
* 141035 153563: contig of 12529 bp in length
* 153564 153663: gap of 100 bp
* 153664 170834: contig of 17171 bp in length.
* Location/Qualifiers
* 1. 170834
* /organism="Homo sapiens"

misc_feature	1. .1607	/mol_type="genomic DNA"
misc_feature	/note="assembly_fragment"	/db_xref="taxon:9606"
misc_feature	1708. .2937	/clone="RP11-66115"
misc_feature	/note="assembly_fragment"	/clone lib="RP11-66115"
misc_feature	3038. .4551	
misc_feature	/note="assembly_fragment"	
misc_feature	4652. .5804	
misc_feature	/note="assembly_fragment"	
misc_feature	5905. .7016	
misc_feature	/note="assembly_fragment"	
misc_feature	7117. .8323	
misc_feature	/note="assembly_fragment"	
misc_feature	8424. .10330	
misc_feature	/note="assembly_fragment"	
misc_feature	10431. .12576	
misc_feature	/note="assembly_fragment"	
misc_feature	12677. .15039	
misc_feature	/note="assembly_fragment"	
misc_feature	15140. .17696	
misc_feature	/note="assembly_fragment"	
misc_feature	17797. .20417	
misc_feature	/note="assembly_fragment"	
misc_feature	20518. .23021	
misc_feature	/note="assembly_fragment"	
misc_feature	23122. .25657	
misc_feature	/note="assembly_fragment"	
misc_feature	25758. .29388	
misc_feature	/note="assembly_fragment"	
misc_feature	29489. .34207	
misc_feature	/note="assembly_fragment"	
misc_feature	34308. .38295	
misc_feature	/note="assembly_fragment"	
misc_feature	38396. .42889	
misc_feature	/note="assembly_fragment"	
misc_feature	42990. .47895	
misc_feature	/note="assembly_fragment"	
misc_feature	47996. .54590	
misc_feature	/note="assembly_fragment"	
misc_feature	54691. .59447	
misc_feature	/note="assembly_fragment"	
misc_feature	clone_end:SP6	
misc_feature	vector_side:left"	
misc_feature	59548. .65836	
misc_feature	/note="assembly_fragment"	
misc_feature	65937. .72591	
misc_feature	/note="assembly_fragment"	
misc_feature	clone_end:T7	
misc_feature	vector_side:left"	
misc_feature	72692. .83266	
misc_feature	/note="assembly_fragment"	
misc_feature	83367. .94803	
misc_feature	/note="assembly_fragment"	
misc_feature	94904. .110268	
misc_feature	/note="assembly_fragment"	
misc_feature	110369. .124647	
misc_feature	/note="assembly_fragment"	
misc_feature	124748. .140934	
misc_feature	/note="assembly_fragment"	
misc_feature	141035. .153563	
misc_feature	/note="assembly_fragment"	
misc_feature	153664. .170834	
misc_feature	/note="assembly_fragment"	

ORIGIN

Query Match	55.1%;	Score 1103.2;	DB 2;	Length 170834;
Best Local Similarity	79.4%;	Pred. No. 1.2e-229;		
Matches 1465;	Conservative 0;	Mismatches 198;	Indels 183;	Gaps 6;
QY	245	CAGAGTTGGAATAATGATGCTGTAGATGCTGGAAAGTATACCATTTGGCTTGGCCACAG	304	

Db	120442	CAAGTTGGAAAAAGTATGTTGATATAGATGCTGGAAGTATACCATTTGGCTTGGGCCAGG	120383
QY	305	CCAAGATGGGCTTCTGCACAGATAGAGAAGATATTAACTCTTTGCATGACGTGTGTTTC	364
Db	120382	CCAAGATGGGCTCTGCACACATGAGAAAGATCAACTCTCTTTGCATGCTGTGTTTC	120323
QY	365	AGAATCTTATGAGAGAAAATAACCTTTCTTATGATTGCATTGGCGGCTGAAAGTTGGAA	424
Db	120322	AGAATCTTATGAGAGAAAATACTTTCTTACGATTGCACTGGCCGACTGGAAAGTTGGAA	120263
QY	425	CAGAGACAATCATCCGACAAATCAAAAGTCTGTGAAGACTTAATTGATGCAGCTGTTGAAG	484
Db	120262	CAGAGACAATCAAGGACAAATCAAAAGTCAAGTGAAGACTTAATTGATGCAGCTGTTGAAG	120203
QY	485	AGTCTGGGAATACAGATATAGAAGGAATCGACACAACTAAATGCATGCTATGGAGCGACAG	544
Db	120202	AGTCTGGGAATATAGATATAGAAGGAACCGACACAACTCATGCTATGGAGCGCCAG	120143
QY	545	CTGCTGTCTTCAATGCTGTAACTGGAATTGAGTCCAGCTTTGGAT-----	591
Db	120142	CTGCTGTCTTCAAGTGTGTTAACTGGATTGAGTCCAGCTCTGGATGACAGTTTGCCC	120083
QY	592	-----	591
Db	120082	TGAGATTGTAGAGAGATATTGCTGTATATCCACAGGAAAAAGTTAGACTGACAGGTAGAG	120023
QY	592	-----GGGCTTC	598
Db	120022	TCAGAGCAGTAGCTCTGCTAATTGAGCCAAATGCAACTTTAATTTTGAACGAGGGCTTC	119963
QY	599	GTGGGACACATATGCAACATGCCCTATGATTTTACAAAGCCTGATATGCTATCTGAATATC	658
Db	119962	GTGGGACACATATGCAACGTCGCTTATGAGCCTGATATGCTCTCTGAATGTC	119903
QY	659	CTATAGTAGATGGAATACTCTCCATACAGTGTCTACCTCAGTGCATTGACCGCTGCTACT	718
Db	119902	CTATAGTAGATGGAATAACCTCCACACAGTGTCTCAGTGCATTGATCACTGCTATTC	119843
QY	719	CTGTCTACTGCAAAAAAGATCCATGCGCCAGTGGCAGAAAAGAGGAATGATTAAGATTTTA	778
Db	119842	CTGTCTACTGCAAAAAAGATCCGTCGCCAGTGGCAGAAAAGAGGAATGATTAAGATTTTA	119783
QY	779	CCTTGAATGATTTTGGCTTCATGATCTTTCACTCACCATTAATGTAACTGGTTCAGAAAT	838
Db	119782	CCTTGAATGATTTTGGCTTCATGATCTTTCACTCACCATTAATGTAACTGGTTCAGTTAT	119723
QY	839	CTCTAGCTCGGATGTGCTGTAATGACTCTCTTAATGACCAAGATAGAGATTAATAATAGTA	898
Db	119722	CTCTAGCTCGGATGTGCTGGAATGACTCTCTTAATGATCAGAAACAGAGATTAATAATAGTG	119663
QY	899	TCTATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATTAGAG	958
Db	119662	TCTATAGTGGCCTGGAAGCCTTTGGGAATGTTAAATTAGAAGACATCTATTTTGATTACAG	119603
QY	959	ATGTGGAGAAGCAATTTATGAAGGCTAGCTCTGAACCTTCAGTCAGAAAACAAAAGGCAT	1018
Db	119602	ATGTGGAGAAGCAATTTATGAAGGAATGTTCCGAACCTTCATCAGAAAACAAAAGGCAT	119543
QY	1019	CTTTACTTGTATCAAAATCAAAATGGAATATGTAGACATCTTCAGTATATGTTCCCTTG	1078
Db	119542	CTTTGTC-TGTATCCAGTGAATAATGGAATATCGTACATCGTCAGTATATGTTCCCC-G	119485
QY	1079	CATCTGTTCTAGCACAGTACTACCTCAGCAATTAGCAGGGAAGAGAAATTGAGTGTTTT	1138
Db	119484	CAACTGTCTCCAGCGCAGTACTACCTCAGCAATGATCAGGGAAGAG-ACTGGAAGTGTCT	119426
QY	1139	CTTATGCTTCTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAGATGCTACAC	1198
Db	119425	CGTACGGTTCTGTGTTTCAGCCGCCACTGTACTCTCTTAAAGTTCACACAGATGCCACAC	119366
QY	1199	CGGGTCTGCTCTTGATTAATAACGCAAGTTTATGTGATCTTAAATCAAGGCTTGATT	1258

Db 119365 CATGGCTCTGCTGTTGATMAAAATAACAGCAAGCTTATGGGATCTTAATCAAGGTGTAGCT 119306
QY 1259 CAAGAACTGGTGTGGCACCAGATGCTTCTGCTGMAAAACATGAAGCTCAGAGAGACACCC 1318
Db 119305 CAAGAAGTTGTGTGACCAAGATGCTTGTGMAAAACATGAAGCTCAGAGAGACACCC 119246
QY 1319 ATCATTTGGTCAACTATATTCGCCAGGGTTCATATAGATTCACTCTTTGAAGGAAGTGTG 1378
Db 119245 ATCACTTGGTCAACTCTGTCCCGGGGCCCAAGAGATTCACTCTTTGAAGAAACATGTT 119186
QY 1379 ACTTAGTTAGGGGTGGATGAAMAAGCAGAGAAGAACTTACGTCGGCGCTCCCACTCCAATG 1438
Db 119185 ACTTAGCTATGGTGGGTGAAMAAGCAGAGAAGAACTTACCTCCGTCTCTCCAATA 119126
QY 1439 ATGACACTTTGGATGAAGGAGTAGGAACTTGTGCATTCAAAACATAGCAACTGAGCATATTC 1498
Db 119125 ATGACACTTGGGATGAAGGAGTAGGAACTTGTGCATTCAAAACATAGTAATGATCATATTC 119066
QY 1499 CAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGAAACCTGMAAGCAGCTG 1558
Db 119065 CAAGCCCTGCTAAGACAGTGCACAAAGACTCCCTGCCACAGCAGAAACGGAGCAGCTG 119006
QY 1559 TCATTAGTAATGGGAACATTAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGG 1618
Db 119005 TCATTAGTAATGGGAACATTAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGG 118946
QY 1619 GCATGGGGTGGGGTATGGGAACATTTGGAAGAATGGGATATCTGGGGATATTTAAAG 1678
Db 118945 G-----GATGG 118939
QY 1679 GATTACATGTTATGTAATTTTATGTGACTGACATGAGCCTGGATGACTATCGTGTAC 1738
Db 118938 GATTACATGTTGCTCAGATTTTATGTGAGTGACATGAGCCTGGATGACCATCGTGTAC 118879
QY 1739 TTGGGAAAGTCTCT-TTGCTCTAATTGCTGACATGCTTCTGTGTGCTGGCCAATGC 1797
Db 118878 TTGGGAAAGTCTCTTTCCTCAATTGTGACATGTTCCCACTGTGTGCTGGCCAATGC 118819
QY 1798 CAAATGTACTCGAATGATGTAAAGGGCTCTGTAAAACTTCAATACCTTTGGCCATTGT 1857
Db 118818 CAAGTGCCTTGAAGTATGTAAAGGGCTCTATAAAACTTCAAGACTCTCTAGCCATTCTG 118759
QY 1858 ATGCATGATGTTGTTGTTTAAACATGGTATATGAATGTGTACTTCTGTCAAGAGAAA 1917
Db 118758 ATACATGAAGTTTAGTTTCAACCATGGTATATGAATCCGTCTCTGTCAAGAGAAA 118699
QY 1918 GCAGAGGTACTAATCTCCAATTAAAAATTTTAAACATGTAAAAA 1963
Db 118698 GCAGAGGTACTAATCTTCAATTAAAAATTTTAAAAACAGATAAGAA 118653

Search completed: June 24, 2004, 09:40:21
Job time : 7619 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2004, 12:56:26 ; Search time 44 Seconds
(without alignments)
565.671 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MFGSLPLNACWPKVGVIV.....PRLPATTAEPBAVISNGEH 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2480	98.8	520	1	HMCS_HUMAN Q01581 homo sapien
2	2362	94.1	520	1	HMCS_RAT P17425 rattus norv
3	2352	93.7	520	1	HMCS_CRIGR P13704 cricetus
4	2097	83.5	522	1	HMCS_CHICK P23228 gallus gall
5	1513	60.3	508	1	HMCM_HUMAN P54868 homo sapien
6	1494	59.5	508	1	HMCM_RAT P22791 rattus norv
7	1493	59.5	480	1	HMCM_MOUSE P54869 mus musculu
8	1466	58.4	508	1	HMCM_PIG O02734 sus scrofa
9	1352.5	53.9	453	1	HMCI_BLAG P54961 blattella g
10	1272.5	50.7	455	1	HMCM_MOUSE P54870 blattella g
11	980	39.0	461	1	HMCS_YEAST P54873 arabidopsis
12	950.5	37.9	491	1	HMCS_YEAST P54839 saccharomyc
13	900	35.8	447	1	HMCS_SCHPO P54874 schizosacch
14	705	28.1	462	1	HMCS_CABEL P54871 caenorhabdi
15	249	9.9	163	1	HMCS_DICDI P54872 dictyosteli
16	152	6.1	350	1	YD79_METKA Q8tv10 methanopyru
17	147	5.9	345	1	PKSG_BACSU P40830 bacillus su
18	145.5	5.8	346	1	Y792_METTH O26883 methanobact
19	144	5.7	350	1	Y677_PYRHO O58410 pyrococcus
20	138.5	5.5	350	1	Y972_PYRFU O51798 pyrococcus
21	135	5.4	349	1	Y871_METMA Q8pyj0 methanosaar
22	132.5	5.3	343	1	Y015_ARCFU O30256 archaeoglob
23	130	5.2	350	1	YL85_PYRAE Q8zvp4 pyrobaculum
24	127	5.1	349	1	Y4E1_METAC Q8tiv0 methanosaar
25	126.5	5.0	350	1	YD69_PYRAB Q9uyy8 pyrococcus
26	123.5	4.9	350	1	BGAL_THETU P26257 thermobacter
27	123	4.9	351	1	Y132_THETU Q97c99 thermoplasma
28	122	4.9	345	1	YF46_METJA Q58941 methanococc
29	122	4.9	351	1	YB55_THETAC Q9hi87 thermoplasma
30	121	4.8	580	1	Y056_MOUSE Q9d621 mus musculu
31	118.5	4.7	596	1	YD49_SULTO O00567 homo sapien
32	116	4.6	348	1	YD49_SULTO Q97lk8 sulfolobus
33	107.5	4.3	776	1	TOPI_RICPR Q9zdk2 rickettsia

34	106.5	4.2	1517	1	GLTB_ECOLI P09831 escherichia
35	106	4.2	367	1	YQI2_CABEL Q09291 caenorhabdi
36	105	4.2	412	1	AATC_MOUSE P05201 mus musculu
37	105	4.2	3358	1	PGCV_MOUSE Q62059 mus musculu
38	103	4.1	489	1	YSVS_CABEL Q10011 caenorhabdi
39	103	4.1	1197	1	DPOM_PODAN Q01529 podospora a
40	102	4.1	360	1	HIS8_LACLA Q02135 lactococcus
41	102	4.1	1163	1	COAA_BACTF Q9x597 bacillus th
42	101	4.0	365	1	Y173_AERPE Q9yas0 aeropyrum p
43	100.5	4.0	776	1	TOPI_RICCN Q92ih1 rickettsia
44	100	4.0	1136	1	C4BA_BACTI P05519 bacillus th
45	99.5	4.0	494	1	ILVC_VIBPA Q87tn4 vibrio para

ALIGNMENTS

RESULT 1	HMCS_HUMAN	STANDARD;	PRT;	520 AA.
AC	Q01581;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).			
GN	HMGCS1 OR HMGCS.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fibroblast;			
RX	MEDLINE=93041939; PubMed=1358203;			
RA	Russ A.P., Ruzicka V., Maerz W., Appelhans H., Gross W.;			
RT	"Amplification and direct sequencing of a cDNA encoding human			
RT	cytosolic 3-hydroxy-3-methylglutaryl-coenzyme A synthase.";			
RL	Biochim. Biophys. Acta 1132:329-331(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A., CHARACTERIZATION, AND MUTAGENESIS.			
RC	TISSUE=Fetal adrenal gland;			
RX	MEDLINE=94304197; PubMed=7913309;			
RA	Rokosz L.L., Boulton D.A., Buckiewicz E.A., Sanyal G., Cueto M.A.,			
RA	Lachance P.A., Hermes J.D.;			
RT	"Human cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase:			
RT	expression, purification, and characterization of recombinant			
RT	wild-type and Cys129 mutant enzymes.";			
RL	Arch. Biochem. Biophys. 312:1-13(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lung;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,			
RA	Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			


```
CC -1- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
CC SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOIDS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66435; CAA47061.1; -.
DR EMBL; L25798; AAA62411.1; -.
DR EMBL; BC000297; AAH00297.1; -.
DR PIR; S27197; S27197.
DR PIR; S45497; S45497.
DR Genew; HGNC:5007; HMGCS1.
DR MIM; 142940; -.
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005625; C:soluble fraction; TAS.
DR GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; TAS.
DR InterPro; IPR000590; HMG_CoA_synth_AS.
DR InterPro; IPR008260; HMG_CoA_synth.
DR Pfam; PF01154; HMG_CoA_synth.1.
DR PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW Transferase; Cholesterol biosynthesis; Multigene family.
FT ACT_SITE 129 129
FT MUTAGEN 129 129 C->A,S: LOSS OF ACTIVITY.
FT CONFLICT 248 248 G->A (IN REF. 1).
FT CONFLICT 251 251 K->N (IN REF. 1).
FT CONFLICT 299 299 E->K (IN REF. 1).
FT CONFLICT 364 364 Q->H (IN REF. 1).
FT CONFLICT 420 420 P->Q (IN REF. 1).
FT CONFLICT 519 520 EH->VW (IN REF. 1).
SQ SEQUENCE 520 AA; 57293 MW; C669212BF86CF9B CRC64;

Query Match 98.8%; Score 2480; DB 1; Length 520;
Best Local Similarity 91.9%; Pred. No. 2.3e-173;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MFGSLPLNAAECWPKDVGIVALEIYFPPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60
DB 1 MFGSLPLNAAECWPKDVGIVALEIYFPPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60

QY 61 DREDINSLCMTYVQNLMERNNISYDCIGRLEVGTEITIIDKSKSVKTNLMQLFEESGNTDI 120
DB 61 DREDINSLCMTYVQNLMERNNISYDCIGRLEVGTEITIIDKSKSVKTNLMQLFEESGNTDI 120

QY 121 EGIDTTNACYGGTAAVFNANVMISSSWD----- 149
DB 121 EGIDTTNACYGGTAAVFNANVMISSSWDGRVALVAAGDIAYATGNARPTGGVCAVALL 180

QY 150 -----GLRGTHMOHAYDFYKPKDMLSEYPIVDGKLSIOCYLSALDRCSVYCKI 198
DB 181 IGPNAPLIFERGLRGTHMOHAYDFYKPKDMLSEYPIVDGKLSIOCYLSALDRCSVYCKI 240

QY 199 HAQMOKEGNDKDFTLNDFGFMTFHSPPYCKLVOKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
DB 241 HAQMOKEGNDKDFTLNDFGFMTFHSPPYCKLVOKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

QY 259 FGDVKLEDYFDRDVEKAFMKASSELFSQKTRASLLVSNQNGNMYTSSVYGSILASVLAQY 318
DB 301 FGDVKLEDYFDRDVEKAFMKASSELFSQKTRASLLVSNQNGNMYTSSVYGSILASVLAQY 360

QY 319 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
DB 361 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420
```

```
QY 379 DYFAENMKLRDPTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRYARRPTPNDDTLDEG 438
DB 421 DYFAENMKLRDPTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRYARRPTPNDDTLDEG 480

QY 439 VGLVHSNIATEHTPSPAKKVPRLPATAAEPAAVISINGEH 478
DB 481 VGLVHSNIATEHTPSPAKKVPRLPATAAEPAAVISINGEH 520

RESULT 2
HMC5_RAT
ID HMC5_RAT STANDARD; PRT; 520 AA.
AC P17425;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 2.3.3.10) (HMG-CoA
DE synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN HMGCS1 OR HMGCS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=90301491; PubMed=1972979;
RA Ayte J., Gil-Gomez G., Hegardt F.G.;
RT "Nucleotide sequence of a rat liver cDNA encoding the cytosolic 3-
RT hydroxy-3-methylglutaryl coenzyme A synthase.";
RL Nucleic Acids Res. 18:3642-3642(1990).
CC -1- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
CC SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOIDS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52625; CAA36852.1; -.
DR PIR; S12736; S12736.
DR InterPro; IPR000590; HMG_CoA_synth_AS.
DR InterPro; IPR008260; HMG_CoA_synth.
DR Pfam; PF01154; HMG_CoA_synth.1.
DR PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW Transferase; Cholesterol biosynthesis; Multigene family.
FT ACT_SITE 129 129 POTENTIAL.
SQ SEQUENCE 520 AA; 57433 MW; CB213A27B0C177CB CRC64;

Query Match 94.1%; Score 2362; DB 1; Length 520;
Best Local Similarity 87.5%; Pred. No. 9.2e-165;
Matches 455; Conservative 11; Mismatches 12; Indels 42; Gaps 1;

QY 1 MFGSLPLNAAECWPKDVGIVALEIYFPPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60
DB 1 MFGSLPLNAAECWPKDVGIVALEIYFPPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60

QY 61 DREDINSLCMTYVQNLMERNNISYDCIGRLEVGTEITIIDKSKSVKTNLMQLFEESGNTDI 120
DB 61 DREDINSLCMTYVQNLMERNNISYDCIGRLEVGTEITIIDKSKSVKTNLMQLFEESGNTDI 120

QY 121 EGIDTTNACYGGTAAVFNANVMISSSWD----- 149
DB 121 EGIDTTNACYGGTAAVFNANVMISSSWD----- 149
```

Db 121 EGIDTTNACYGSTA AAVNAVNWIESSWDGRYALVVAGDIAIYASGNARPTGGVAVALL 180
QY 150 -----GLRGTHMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKKI 198
Db 181 IGPNAVPIFDRGLRGTHMOHAYDFYKPDMLSEYPVVDGKLSIQCYLSALDRCSYVCKKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 258
Db 241 RAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLVSNQNGNMYTSSVYGSGLASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASAELEFNQKTKASLVSNQNGNMYTSSVYGSGLASVLAQY 360
QY 319 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSRTGVAP 378
Db 361 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSRTGVAP 420
QY 379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKGRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKGRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNTATEHIPSPAKKVPRLPATSGEPESAVISNGEH 520

RESULT 3
HMCS_CRIGR STANDARD; PRT; 520 AA.
ID HMCS_CRIGR
AC P13704;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN HMGCS1 OR HMGCS.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=86140166; Pubmed=2869035;
RA Gil G., Goldstein J.L., Slaughter C.A., Brown M.S.,
RT "Cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase from the hamster. I. Isolation and sequencing of a full-length cDNA.",
RT J. Biol. Chem. 261:3710-3716(1986).
RL J. Biol. Chem. 261:3710-3716(1986).
CC -1- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: PRODUCTION OF MEVLONATE FROM HMG-CoA PRIOR TO THE SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOIDS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L00334; AAA37076.1; -;
DR EMBL; L00326; AAA37076.1; JOINED.
DR EMBL; L00327; AAA37076.1; JOINED.
DR EMBL; L00328; AAA37076.1; JOINED.
DR EMBL; L00329; AAA37076.1; JOINED.
DR EMBL; L00330; AAA37076.1; JOINED.

DR EMBL; L00331; AAA37076.1; JOINED.
DR EMBL; L00332; AAA37076.1; JOINED.
DR EMBL; L00333; AAA37076.1; JOINED.
DR PIR; A25332; A25332.
DR InterPro; IPR000590; HMG-CoA synth AS.
DR Pfam; PF01154; HMG-CoA synth; 1.
DR PROSITE; PS01226; HMG-CoA SYNTHASE; 1.
KW Transferase; Cholesterol biosynthesis; Multigene family.
FT ACT SITE 129 129 POTENTIAL.
SQ SEQUENCE 520 AA; 57318 MW; 2D4CAAE7DEE5D6B CRC64;

Query Match 93.7%; Score 2352; DB 1; Length 520;
Best Local Similarity 87.5%; Pred. No. 5e-164;
Matches 455; Conservative 9; Mismatches 14; Indels 42; Gaps 1;

QY 1 MGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKTYTIGLQAKMGFCT 60
Db 1 MGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKTYTIGLQAKMGFCT 60
QY 61 DREDINSLCMTVVQNMERNNSLYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNMERNNSLYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY 121 EGIDTTNACYGSTA AAVNAVNWIESSWD----- 149
Db 121 EGIDTTNACYGSTA AAVNAVNWIESSWDGRYALVVAGDIAIYATGNARPTGGVAVALL 180
QY 150 -----GLRGTHMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKKI 198
Db 181 IGPNAVPIFDRGLRGTHMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 258
Db 241 RAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLVSNQNGNMYTSSVYGSGLASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASAELEFNQKTKASLVSNQNGNMYTSSVYGSGLASVLAQY 360
QY 319 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSRTGVAP 378
Db 361 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSRTGVAP 420
QY 379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKGRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKGRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNTATEHIPSPAKKVPRLPATAESESAVISNGEH 520

RESULT 4
HMCS_CHICK STANDARD; PRT; 522 AA.
ID HMCS_CHICK
AC P23228;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN HMGCS1 OR HMGCS.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91112772; Pubmed=1980405;
RA Kattar-Coolley P.A., Wang H.-H.L., Mende-Mueller L.M., Mizioro H.M.;

```
RT "Avian liver 3-hydroxy-3-methylglutaryl-CoA synthase: distinct genes
RT encode the cholesterologenic and ketogenic isozymes.";
RL Arch. Biochem. Biophys. 283:523-529(1990).
RN [2]
RP CHARACTERIZATION, AND MUTAGENESIS.
RX MEDLINE=95210320; Pubmed=7696316;
RA Misra I., Charlier H.A. Jr., Miziorko H.M.;
RT "Avian cytosolic 3-hydroxy-3-methylglutaryl-CoA synthase: evaluation
RL Biochim. Biophys. Acta 1247:253-259(1995).
CC -1- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
CC SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOLIDS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60657; AAA62737.1; -.
DR InterPro; IPR000590; HMG_COA_synt_AS.
DR InterPro; IPR008260; HMG_COA_synth.
DR Pfam; PF01154; HMG_COA_synt; 1.
DR PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW Transferase; Cholesterol biosynthesis; Multigene family.
FT ACT_SITE 129 129 POTENTIAL.
SQ SEQUENCE 522 AA; 57559 MW; BFF7947C3E963C4C CRC64;

Query Match 83.5%; Score 2097; DB 1; Length 522;
Best Local Similarity 76.6%; Pred. No. 2e-145;
Matches 400; Conservative 36; Mismatches 42; Indels 44; Gaps 3;

QY 1 MPGSLPLNBEACWPKDVGIVALEIYFPPSQYVDQAELEKYDVGADGKYTIGLGOAKMGFCT 60
Db 1 MPGSLPVNTESCWPKDVGIVALEIYFPPSQYVDQTELEKYDVGADGKYTIGLGOAKMGFCS 60
QY 61 DREDINSLCMTVVQNLERNNLSDYDCIGRLEVGTETIIDKSKSVKTNLMQLEESGNTDI 120
Db 61 DREDINSLCMTVVQNLERNNLSDYDCIGRLEVGTETIIDKSKSVKTNLMQLEESGNTDV 120
QY 121 EGIDTNTACGCGTAAVENAVNWIESSWD----- 149
Db 121 EGIDTNTACGCGTAALFNALNWIESSWDGRYALVAGDIAYATGNARPTGAGAVAML 180
QY 150 -----GLRGTMHQAHYDFYKPKDMLSEYPIVDGKSTQCYSALDRCSVYCKKI 198
Db 181 VGSNAPLIFERGLRGTMHQAHYDFYKPKDMSGEYVVDGKSTQCYSALDRCSVYRNKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRD-KNSIYSGLE 257
Db 241 HAQWQKEGTDRGFLLNDFGFMIFHSPYCKLVOKSVARLLNDFLSDQNAETANGVFSGLE 300
QY 258 AFGDVKLEDTYFDRDVEKAFMKASSELFSOKTKASLTVSNONGNMYYTSSVYGSILASVLAQ 317
Db 301 AFRDVKLEDTYFDRDVEKAFMKASAELEFNQTKASLTVSNONGNMYYTPSVYGGCLASILAQ 360
QY 318 YSPQOLAGKRIGVFSYSGSLAATLYSLKVTODATPGSALDKITASLCDLKSRIDSRITGVA 377
Db 361 YSPEHLAGQRISSEFSYSGSFAATLYSIRVTDATPGSALDKITASLSDLKARIDSRKICIA 420
QY 378 PDVFAENMKLRBDTHLVNVIYFGSISDLFEGTWYLVVRVDEKHRRRTYARRPTPNDDTLDE 437
Db 421 PDVFAENMKIRQETHLVNVIYFGSIVEDLFEGTWYLVVRVDEKHRRRTYARRPVNGDPLEA 480
QY 438 GVGLVHNSINATEHIPSAPKAVPRLPATA-AEPEAAVISNGEH 478
```

```
Db 481 GVEVHPGIVHEHIPSAPKAVPRIPATESEGVTAISNGVH 522

RESULT 5
HMCN HUMAN STANDARD; PRT; 508 AA.
ID _HMCN_HUMAN
AC P54868;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor
DE (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme
DE A synthase).
GN HMGCS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95200282; Pubmed=7893153;
RA Mascaro C., Buesa C., Ortiz J.A., Haro D., Hegardt F.G.;
RT "Molecular cloning and tissue expression of human mitochondrial 3-
RT hydroxy-3-methylglutaryl-CoA synthase.";
RL Arch. Biochem. Biophys. 317:385-390(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97449286; Pubmed=9305755;
RA Boukaftane Y., Mitchell G.A.;
RT "Cloning and characterization of the human mitochondrial 3-hydroxy-3-
RL Gene 195:121-126(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; Pubmed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marasina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton A., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 36-508 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95154824; Pubmed=7851882;
RA Boukaftane Y., Duncan A., Wang S., Labuda D., Robert M.-F.,
RA Sarrazin J., Schappert K., Mitchell G.A.;
RT "Human mitochondrial HMG CoA synthase: liver cDNA and partial genomic
RT cloning, chromosome mapping to 1p12-p13, and possible role in
RT vertebrate evolution.";
RL Genomics 23:552-559(1994).
CC -1- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-COA LYASE FOR KETONE
```



```
CC BODY BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- TISSUE SPECIFICITY: Liver and kidney.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83618; CAA58593.1; -.
DR EMBL; U81859; AAB72036.1; -.
DR EMBL; U81851; AAB72036.1; JOINED.
DR EMBL; U81852; AAB72036.1; JOINED.
DR EMBL; U81853; AAB72036.1; JOINED.
DR EMBL; U81854; AAB72036.1; JOINED.
DR EMBL; U81855; AAB72036.1; JOINED.
DR EMBL; U81856; AAB72036.1; JOINED.
DR EMBL; U81857; AAB72036.1; JOINED.
DR EMBL; U81858; AAB72036.1; JOINED.
DR EMBL; BC044217; AAH44217.1; -.
DR EMBL; U12788; AAA92673.1; -.
DR EMBL; U12789; AAA92674.1; -.
DR PIR; S71623; S71623.
DR Genew; HGNC:5008; HMGCS2.
DR MIM; 600234; -.
DR GO; GO:0005739; C:mitochondrion; TAS.
DR InterPro; IPR000590; HMG_CoA_synt_AS.
DR InterPro; IPR008260; HMG_CoA_synth.
DR Pfam; PF01154; HMG_CoA_synt; 1.
DR PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW Transferase; Cholesterol biosynthesis; Mitochondrion; Transit peptide;
KW Multigene family.
FT TRANSIT 1 37 MITOCHONDRION (PROBABLE).
FT CHAIN 38 508 HYDROXYMETHYLGUTARYL-COA SYNTHASE.
FT ACT SITE 166 166 POTENTIAL.
SQ SEQUENCE 508 AA; 56635 MW; BD362D631F7C3C80 CRC64;

Query Match 60.3%; Score 1513; DB 1; Length 508;
Best Local Similarity 60.4%; Pred. No. 7.5e-103;
Matches 282; Conservative 69; Mismatches 74; Indels 42; Gaps 1;

QY 4 SLPLNAEACWPKDVGIVALEIYFPPSQYVDQAELEKYDGDAGKYTIGLGOAKMGFCTDRE 63
DB 41 AVPLAKDTWPKDVGILALEVYFPAQYVDQTDLEKYNVNEAGKYTVGLGQTRMGFCSVQE 100
QY 64 DINSLCMTVVQNLMERNNLSDYDCIGRLVEGTETIIDKSKSVKTNLMQLFEESGNTDIEGI 123
DB 101 DINSLCLTVVQRLMERIQLPWDSVGRLEVGTEETIIDKSKAVKTVLMELFQDSGNTDIEGI 160
QY 124 DTTNACYGCGTAAVFNAVNMTIESSSWD----- 149
DB 161 DTTNACYGCGTASLFNAAANWMESSWDGRYAMVVCGDIAVYPSGNARPTGGAGAVAMLIGP 220
QY 150 -----GLRGTHMQHAYPFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQ 201
DB 221 KAPLALERGLRGTHMENYVDFYKPNLASEYPIVDGKLSIQCYLRALDRCYTSYRKKIIONQ 280
QY 202 WQKEGNDKDFTLNDFGFMIHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEAFGD 261
DB 281 WKQAGSDRPFLLDLDLQYMFTHFPCKMVQKSLARLMFNDFLSASSDTQTSLYKGLEAFGG 340
QY 262 VKLEDYTFDRDVEKAFMKASSELFSQKTAKSLVSNONGNMYTSSVYGSILASVLAQYSPQ 321
DB 341 LKLEDYTNKDLKALKASQDMFDKKTAKSLYLSTHNGNMYTSSLYGLASLSHSAQ 400
QY 322 QLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAPDYF 381
DB 401 ELAASRIGAFSYSGSLAASFFSFRVSDAAPGSPLDKLVSTSDLPKRLASRKCVSEEF 460
```

```
QY 382 AENMKLRDTHLVNIPOGSIDSLFEGTWYLVVDEKRRRTYARRP 428
DB 461 TEIMNOREQFHYHKVNSPPGDTNSLFGITWYLERVDEQHRKRYARRP 507

RESULT 6
HMCN_RAT
ID HMCN_RAT STANDARD; PRT; 508 AA.
AC P22791;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor
DE (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme
DE A synthase).
GN HMGCS2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=90251660; PubMed=19711108;
RA Ayte J., Gil-Gomez G., Haro D., Marrero P.F., Hegardt F.G.;
RT "Rat mitochondrial and cytosolic 3-hydroxy-3-methylglutaryl-CoA
RT synthases are encoded by two different genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:3874-3878(1990).
CC -1- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-COA LYASE FOR KETONE
CC BODY BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- TISSUE SPECIFICITY: Liver and kidney.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33648; AAA41336.1; -.
DR PIR; A35865; A35865.
DR InterPro; IPR000590; HMG_CoA_synt_AS.
DR InterPro; IPR008260; HMG_CoA_synth.
DR Pfam; PF01154; HMG_CoA_synt; 1.
DR PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW Transferase; Cholesterol biosynthesis; Mitochondrion; Transit peptide;
KW Multigene family.
FT TRANSIT 1 37 MITOCHONDRION (PROBABLE).
FT CHAIN 38 508 HYDROXYMETHYLGUTARYL-COA SYNTHASE.
FT ACT SITE 166 166 POTENTIAL.
SQ SEQUENCE 508 AA; 56912 MW; EC37693A5541D47E CRC64;

Query Match 59.5%; Score 1494; DB 1; Length 508;
Best Local Similarity 59.1%; Pred. No. 1.8e-101;
Matches 275; Conservative 76; Mismatches 72; Indels 42; Gaps 1;

QY 6 PLNAEACWPKDVGIVALEIYFPPSQYVDQAELEKYDGDAGKYTIGLGOAKMGFCTDREDI 65
DB 43 PLAKDTWPKDVGILALEVYFPAQYVDQTDLEKFNVNEAGKYTVGLGQTRMGFCSVQEDI 102
QY 66 NSLCMTVVQNLMERNNLSDYDCIGRLVEGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDT 125
DB 103 NSLCITVVQRLMERITKLPMDAVGRLEVGTEETIIDKSKAVKTVLMELFQDSGNTDIEGIDT 162
QY 126 TNACYGCGTAAVFNAVNMTIESSSWD----- 149
```

```

Db      163 TNACYGCTASLFNANMMESSYWDGRYALVCGDIAVYPSGNRPPTGGAGAVAMLIGPKA 222
Qy      150 -----GLRGTMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKKIHAQWQ 203
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      223 PLVLEQGLRGTHMENAYDFYKPNLASEYPLVDGKLSIQCYLRALDRCYAAYRRRIQNWK 282
Qy      204 KEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEAFGDVK 263
          :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      283 QAGNNQPFLLDVQYMI FHTPFCKMVQKSLARLMFNDFLSSSSDKQNNLLYKGLFAFKGLK 342
Qy      264 LEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQYSPQOL 323
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      343 LEETYNKDVDPKALLKASLDMEFNKTKASLYLSTNNGNMYTSSLYGCLASLSHSAQEL 402
Qy      324 AGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSRTGVADVFAE 383
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      403 AGSRIGAFSYSGSLAASFSPFRVSKDASPSPLEKLVSSVSDLPKRLDSRRRMSPEEFTE 462
Qy      384 NMKLRDTHLVNYIPOGSIDSLFEGTWYLVRVDEKHRRTYARRP 428
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      463 IMNQREQFYHKVNFSPPGDTSNLFPGTWYLERVDEMHRRKYARRP 507

```

RESULT 7

HMCN_MOUSE STANDARD; PRT; 480 AA.

```

ID _HMCN_MOUSE Q64740;
AC P54869; Q64740;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor
DE (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme
A synthase) (Fragment).
GN HMGCS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHS; TISSUE=Liver;
RX MEDLINE=95154824; PubMed=7851882;
RA Boukafane Y., Duncan A., Wang S., Labuda D., Robert M.-F.,
RT "Human mitochondrial HMG CoA synthase: liver cDNA and partial genomic
RT cloning, chromosome mapping to 1p12-p13, and possible role in
RT vertebrate evolution."
RT Genomics 23:552-559(1994).
RL -1- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-COA LYASE FOR KETONE
CC BODY BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- TISSUE SPECIFICITY: Liver and kidney.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U12790; AAA92675.1; -.
CC EMBL; U12791; AAA92676.1; -.
CC SWISS-2DPAGE; P54869; MOUSE.
CC MGD; MGI:101939; Hmgcs2.
CC InterPro; IPR000590; HMG_COA_synth_AS.
CC InterPro; IPR008260; HMG_COA_synth.
CC Pfam; PF01154; HMG_Coa_synth; 1.

```

```

DR PROSITE; PS01226; HMG COA SYNTHASE; 1.
KW Transferase; Cholesterol Biosynthesis; Mitochondrion; Transit peptide;
KW Multigene family.
FT NON_TER 1 1
FT TRANSIT <1 9 MITOCHONDRION (PROBABLE).
FT CHAIN 10 480 HYDROXYMETHYLGUTARYL-COA SYNTHASE.
FT ACT_SITE 138 138 POTENTIAL.
SQ SEQUENCE 480 AA; 53786 MW; 7AA667A92862214D CRC64;

```

Query Match 59.5%; Score 1493; DB 1; Length 480;
 Best Local Similarity 59.4%; Pred. No. 2e-101;
 Matches 276; Conservative 74; Mismatches 73; Indels 42; Gaps 1;

```

Qy      6 PLNAEACWPKDVGIVALEIYFPSSQYVDQAELEKVDGVDAGKTYTIGQAAMGFCIDREDI 65
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      15 PLAKTDTPKDVGLALEVYFPAQYVDQTDLEKFNVEAGKTYVGLQTRMGFCSVQEDI 74
Qy      66 NSLCMTVVQNLMEERNLLSYDCIGRLEVGTEIIDKSKSVKTNLMQLFEESGNTDIEGIDT 125
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      75 NSLCILTVQRLMERTKLPWDVAVGRLEVGTEIIDKSKAVKTVLMELFQDSGNTDIEGIDT 134
Qy      126 TNACYGCTAAVFNAVNWIESSWD----- 149
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      135 TNACYGCTASLFNANMMESSYWDGRYALVCGDIAVYPSGNRPPTGGAGAVAMLIGPKA 194
Qy      150 -----GLRGTMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKKIHAQWQ 203
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      195 PLVLEQGLRGTHMENAYDFYKPNLASEYPLVDGKLSIQCYLRALDRCYAAYRRRIQNWK 254
Qy      204 KEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEAFGDVK 263
          :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      255 QAGNNQPFLLDVQYMI FHTPFCKMVQKSLARLMFNDFLSSSSDKQNNLLYKGLFAFKGLK 314
Qy      264 LEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQYSPQOL 323
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      315 LEETYNKDVDPKALLKASLDMEFNQKTKASLYLSTNNGNMYTSSLYGCLASLSHSAQEL 374
Qy      324 AGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSRTGVADVFAE 383
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      375 AGSRIGAFSYSGSLAASFSPFRVSKDASPSPLEKLVSSVSDLPKRLDSRRRMSPEEFTE 434
Qy      384 NMKLRDTHLVNYIPOGSIDSLFEGTWYLVRVDEKHRRTYARRP 428
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      435 IMNQREQFYHKVNFSPPGDTSNLFPGTWYLERVDEMHRRKYARRP 479

```

RESULT 8

HMCN_PIG STANDARD; PRT; 508 AA.

```

ID _HMCN_PIG Q02734;
AC 002734;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor
DE (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme
A synthase).
GN HMGCS2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97293201; PubMed=9164842;
RA Adams S.H., Alho C.S., Asins G., Hegardt F.G., Marrero P.F.;
RT "Gene expression of mitochondrial 3-hydroxy-3-methylglutaryl-CoA
RT synthase in a poorly ketogenic mammal: effect of starvation during
RT the neonatal period of the piglet."
RL Biochem. J. 324:65-73(1997).
CC -1- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-

```

CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-COA LYASE FOR KETONE
CC BODY BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U90884; AAC48727.1; -.
DR InterPro; IPR000590; HMG_CoA_synth_AS.
DR InterPro; IPR008260; HMG_CoA_synth.
DR Pfam; PF01154; HMG_CoA_synth; 1.
DR PROSITE; PS01226; HMG_CoA SYNTHASE; 1.
KW Transferase; Cholesterol biosynthesis; Mitochondrion; Transit peptide;
KW Multigene family.
FT TRANSIT 1 37 MITOCHONDRION (PROBABLE).
FT CHAIN 38 508 HYDROXYMETHYLGUTARYL-COA SYNTHASE.
FT ACT_SITE 166 166 POTENTIAL.
SQ SEQUENCE 508 AA; 56933 MW; 5479DE6F70B3C0F6 CRC64;

Query Match 58.4%; Score 1466; DB 1; Length 508;
Best Local Similarity 59.8%; Pred. No. 2e-99;
Matches 278; Conservative 68; Mismatches 77; Indels 42; Gaps 1;

QY 4 SLPLNAEACWPKDVGIVALEIYFSPQYVDQAELEKYDGDAGKTYTIGLGOAKMGFCTDRE 63
Db 41 AVFPAKADTWPKDVGILALEYFPAQYVDQTDLEKFDNVEAGRYTVGLGQTHMGFSVQE 100

QY 64 DINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMOLFEESENTDIEGI 123
Db 101 DINSCLTVVQRLMERTQLPWSVSWGLEVTETIIDKSKSVKTVLMELFQDSGNTDIEGI 160

QY 124 DTTNACYGCTAAVFNAVNWIESSSWD----- 149
Db 161 DTTNACYGCTASLFNAANWVSSAWDGRYAVVVCGDIAVYPRGNSRPTGAGAVAMLVGP 220

QY 150 -----GLRGTMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKIHAQ 201
Db 221 EAPLALERGLRGTHMENAYDFYKPNATSEYPLVDGKLSIQCYLRALDRCTYLYRQKIEKQ 280

QY 202 WQKEGNDKDFTLNDFGFMIFFHSFYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEAFGD 261
Db 281 WKQAGIERHFTLDDLQFMIFHTPFCKLVQKSLARLMFSDFLIADSDTQSSLYKGLEAFRG 340

QY 262 VKLEDYFDRDVEKAFMKASSELFSQKTAKSLVSNQNGNMYTSSVYGSVLAAQYSPQ 321
Db 341 QKLEDYANKDIEKAFQKASLDLFNKTKPSLYLSLHNGNMYTSSLYGCLASLTSQCSAQ 400

QY 322 QLAGKRIGVFSYSGLAATVYSLKVTDATPGSALDKITASLCLDKSRLDSRTGVA PDVF 381
Db 401 DLASGRIGAFSYSGGLAASFYSLRVSQDASPGSPLEKLVSVSDLPERLASRKRVSPDEF 460

QY 382 AENMKLRBEDTHLVNYPQGSIDSLFEGTWYLVRVDEKHRRTYAR 426
Db 461 TEIMNQREQYHKNFTPPGDPNSLFPGTWYLERVDLYRKYAR 505

RESULT 9
HMC1_BLAG
ID HMC1_BLAG STANDARD; PRT; 453 AA.
AC P54961;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase 1 (EC 2.3.3.10) (HMG-CoA synthase
DE 1) (3-hydroxy-3-methylglutaryl coenzyme A synthase 1).
GN HMGCS-1.

OS Blatella germanica (German cockroach).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blaberoidea;
OC Blattellidae; Blattellinae; Blatella.
OX NCBI_TaxID=6973;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94039108; PubMed=7901012;
RA Martinez-Gonzalez J., Buesa C., Piulachs M.D., Belles X.,
RA Hegardt F.G.;
RT "3-hydroxy-3-methylglutaryl-coenzyme-A synthase from Blatella
RT germanica. Cloning, expression, developmental pattern and tissue
RT expression.";
RL Eur. J. Biochem. 217:691-699(1993).
CC -1- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC form HMG-CoA, which is the substrate for HMG-CoA reductase (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
CC SYNTHESIS OF STEROLS AND ISOPRENOIDS SUCH AS JUVENILE HORMONES.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X73679; CAA52032.1; -.
DR PIR; S38986; S38986.
DR InterPro; IPR000590; HMG_CoA_synth_AS.
DR InterPro; IPR008260; HMG_CoA_synth.
DR Pfam; PF01154; HMG_CoA_synth; 1.
DR PROSITE; PS01226; HMG_CoA SYNTHASE; 1.
KW Transferase; Sterol biosynthesis; Multigene family.
FT ACT_SITE 116 116 POTENTIAL.
SQ SEQUENCE 453 AA; 50332 MW; C79EB2376270F348 CRC64;

Query Match 53.9%; Score 1352.5; DB 1; Length 453;
Best Local Similarity 56.6%; Pred. No. 3.3e-91;
Matches 259; Conservative 66; Mismatches 84; Indels 49; Gaps 4;

QY 13 WPKDVGIVALEIYFSPQYVDQAELEKYDGDAGKTYTIGLGOAKMGFCTDREDINSLCMTV 72
Db 2 WPSDVGIVALELIFPSQYVDQVLEVDNVSAGKYTVGLGQARMGFCTDREDINSLCLTV 61

QY 73 VQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMOLFEESENTDIEGIDTTNACYG 132
Db 62 VSRLMERWSIPYSQIGRLEVGTETLLDKSKSVKTVLMQLFKD--NTDIEGVDTVNACYG 119

QY 133 TAAVFNAVNWIESSSWD-----G 150
Db 120 TSALFNAISWVESSSWDGRYALVVAAGDIAVYAKGSARPTGAGAVAMLVGANAPLVFDRG 179

QY 151 LRGTMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKIHAQWQKEGNDKD 210
Db 180 VRSSHQHAYDFYKPDLSLYPTVDGKLSIQCYLSALDHICYQLYCSKI---QKQLEK- 234

QY 211 FTINDFGFMIFFHSFYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEAFGVDKLEDYFD 270
Db 235 FDIERLDAVLFHAPYCKLVQKSLARLVINDFVRASEERTTKYSSLSLEALKGVKLEDYFD 294

QY 271 RDVEKAFMKASSELFSQKTAKSLVSNQNGNMYTSSVYGSVLAAVLAQYSPQOLAGKRIGV 330
Db 295 REVEKAVMTYSKMFEEKTKPSLLANQVGNMYPSTLYGLVSLVSKSAQELAGKRYAL 354

QY 331 FSYSGLAATVYSLKVTDATPGSALDKITASLCLDKSRLDSRTGVA PDVFAENMKLRD 390
Db 355 FSYSGGLASSMFSLRISSDASAKSSLQRLVSNLSHIKQLDLRHKVSPEEFAQTMEIREH 414


```

CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: Production of mevalonate from HMG-CoA prior to the
CC synthesis of sterols and isoprenoids.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X96617; CA65437.1; -.
DR EMBL; Z50178; CAA90557.1; -.
DR PIR; S58202; S58202.
DR GerMOnline; 142661; -.
DR SGD; S0004595; ERG13.
DR InterPro; IPR000590; HMG_COA_synth_AS.
DR InterPro; IPR008260; HMG_COA_synth.
DR Pfam; PF01154; HMG_COA_synth_1.
DR PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
FW Transferase; Sterol biosynthesis.
KW ACT SITE 159 POTENTIAL.
SQ SEQUENCE 491 AA; 55013 MW; 44DFF3C0B0356D9E CRC64;

Query Match 37.9%; Score 950.5; DB 1; Length 491;
Best Local Similarity 43.4%; Pred. No. 7.7e-62;
Matches 198; Conservative 72; Mismatches 127; Indels 59; Gaps 7;

QY 14 PKDVGIVALEIYFPSPQYVDQAELEKXYDGDVADGAKYTIIGLGQAKMGFCTDRREDINSLCMTVV 73
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 46 PQNVGKIGIQIYIPTQCVNQSELEKFDGVSQCKYTIIGLGQTNMSFVNDREDIYSMSLTVL 105
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 74 QNLMERNNLSYDCIGRLVGTETITIDKSQSVKTNLMQLFESGNTDIEGIDTNACYGGT 133
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 106 SKLIKSYNIDTNKIGRLVGTETLIDKSQSVKSLVMQLFGE--NTDVEGIDTLNACYGGT 163
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 134 AAVFNAVNMWIESSWDG-----LR 152
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 164 NALFNSLNMWIESSNAMDGRDAIVVCGDIAIYDKGAARPTGAGTVAWMIGPDAPIVDSVR 223
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 153 GTHMQHAYDFFYKPPDMISEYPIVDGKLSIOCYLSALDRCYSVYCKIHAQW--QKEGNDK 209
   :::||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 224 ASYMEHAYDFFYKPPDFTSEYPIVDGHFSLTCYKALDQVYKSYSKAISKGLVSDPAGSDA 283
   :::||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 210 DFTLNDGFEMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYGLBA-FGDVYKLEDTY 268
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 284 LNVLYKFPDYNVHFVPTCKLVTKSYGRLLYNDF----RANQLFPEVDALATRDYDESL 338
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 269 FDRDVEKAFMKASSELFSQKTKASLIVSNQNGNMYSVVGSLASVLAQYSPOQLAGKRI 328
   :::||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 339 TDKNIEKTFVNVAKPFHKERVAOSLIIVPTNTGNMYTASVYAAFAFSLNLYVGSDDLQGRKV 398
   :::||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 329 GVSFYSGLAATLYSLKYTDATPGSALDKITASLCLDKSRLDSRTGVAAPDVAEENMKLR 388
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 399 GLFSYSGSLAASLYSCKIVD-----VQHIKEL-DITNKLAKRITETPKDYEAIEELR 451
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 389 EDTTHLVNYYIPQGSIDSLFEGTWYLVVRVDEKRRTY 424
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 452 ENAHLKKNFKPQGSIEHLQSGVYYLLTNIDDKFRRSY 487
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 13
HMC5_SCHPO STANDARD; PRT; 447 AA.
ID_HMC5_SCHPO AC P54874;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase (EC 2.3.3.10) (HMG-CoA synthase)
DE (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN HCS OR SPAC4F8.14C.

```

OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96353436; PubMed=8750242;
RA Katayama S., Adachi N., Takao K., Nakagawa T., Matsuda H.,
RA Kawamukai M.;
RT "Molecular cloning and sequencing of the hcs gene, which encodes 3-
RT hydroxy-3-methylglutaryl coenzyme A synthase of Schizosaccharomyces
RT pombe.";
RL Yeast 11:1533-1537(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woldward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Welljeds I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzom K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadiou E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas R., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -|-
CC Nature 415:871-880(2002).
CC -|-
CC FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC form HMG-CoA, which is the substrate for HMG-CoA reductase (By
CC similarity).
CC -|-
CC CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -|-
CC PATHWAY: Production of mevalonate from HMG-CoA prior to the
CC synthesis of sterols and isoprenoids.
CC -|-
CC SIMILARITY: Belongs to the HMG-CoA synthase family.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC -----
CC EMBL; U32187; AAB17601.1; -;
CC EMBL; Z98530; CAB11060.1; -;
CC PIR; S61875; S61875.
CC
CC GenedB_SPombe; SPAC4F8.14c; -;
CC InterPro; IPR000590; HMG_COA_synth_AS.
CC InterPro; IPR008260; HMG_COA_synth.
CC Pfam; PF01154; HMG_COA_synth; 1.
CC PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
CC
CC Transferase; Sterol biosynthesis.
CC
CC ACT_SITE 118 118 POTENTIAL.
CC
CC SEQUENCE 447 AA; 49239 MW; 919BDDBD9207B886 CRC64;
SQ


```
Query Match          35.8%; Score 900; DB 1; Length 447;
Best Local Similarity 41.5%; Pred No.3.2e-58;
Matches 191; Conservative 72; Mismatches 133; Indels 64; Gaps 6;

QY      15 KDVGIVALEIYFPSQYVDQAELEKYDGDVDAKYYTIGLGQAKMGFCTDREDINSLCMTVQN 74
      ||:|:| | | :|:|:|:| | | :|:| | | | | | | | | | | | | | | | | :
      6 KDIGIKGLVLYTPNQYVEQAALEAHADGVSTGKYTIGLGTQWAFVDDRREDIYSFGLTALS 65

QY      75 NLMERNNLSYDCIGRLVEGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGGA 134
      |:|:| | : | | | | | | | | | | | | | | | | | | | | | | | | | |
      66 QLIKRYQIDISKIGRLVEGTETIIDKSKSVKSLVMQLFGD--NHNVEGIDCVNACYGGVN 123

QY      135 AVFNANVMISSSWD-----GLR 152
      |:|:| | :|:|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      124 ALFNTIDWIESSAWDGRDGIIVAGDIALYAKGNARPTGAGCVALLVGNAPIVFEEGLR 183

QY      153 GTHMQHAYDFYKPDMLSEYPIVDKLSIQCYLSALDRCTSYVCKKIHAQWQKEGNDKFT 212
      ||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
      184 GTYMQHAYDFYKPDLTSEYPIVDGHFSLECYVKALDGAAYANYVRDVA--KNGKSQGLG 240

QY      213 LNDEGFMI FHSPYCKLVQKSLARMLNDFL---NDQNRDKNSIYSGLEAFGVDKLEDTY 268
      |:|:| | :|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      241 LDRFDYCI FHAPTCQVQKAYARLLYTDSAEPSPNELEGVRELLSTLDA-----KKS 294

QY      269 FDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQYSPQQLAGRI 328
      |:|:| | :|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      295 TDKALEKGLMAITKERFNKRVSVPVYAPTNCGNMYTASIFSCLTALLSRVPADELKGRV 354

QY      329 GVFSYGSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAPDVAENMKLR 388
      |:|:| | :|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      355 GAVSYGSGLASFPFSFVVKGDVSE-----IAKKTNLVNDLNRHCLTPTQYEAEIELR 407

QY      389 EDTHHLVNIYIPQGSIDSLFEETWYLVRVDEKRRTYARRP 428
      |:|:| | :|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      408 HQAHLKKNFTPKGSIERLRSCTYYLTGIDMFRRSYSVKP 447

Db

RESULT 14
HMCS_CAEEL STANDARD; PRT; 462 AA.
ID_HMCS_CAEEL
AC P54871; Q22962;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase (EC 2.3.3.10) (HMG-CoA synthase)
DE (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN F25B4.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Gattung S.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 5-462 FROM N.A.
RA Mitchell G.A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC form HMG-CoA, which is the substrate for HMG-CoA reductase (by
CC similarity).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + CoA.
CC -1- PATHWAY: Production of mevalonate from HMG-CoA prior to the
CC synthesis of sterols and isoprenoids.
CC -1- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC -----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U64842; AAB37084.1; -.
DR EMBL; U12787; AAA92672.1; -.
DR PIR; T25726; T25726.
DR WormPep; F25B4.6; CE09624.
DR InterPro; IPR000590; HMG_CoA_synth_AS.
DR InterPro; IPR008260; HMG_CoA_synth.
DR Pfam; PF01154; HMG_CoA_synth_1.
DR PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW Transferrase; Sterol biosynthesis.
FT ACT SITE 124 124 POTENTIAL.
SQ SEQUENCE 462 AA; 51415 MW; 89BDE382588F6D9F CRC64;

Query Match          28.1%; Score 705; DB 1; Length 462;
Best Local Similarity 37.1%; Pred. No.5.7e-44;
Matches 176; Conservative 77; Mismatches 151; Indels 70; Gaps 13;

QY      16 DVGIVALEIYFPSQYVDQAELEKYDGDVDAKYYTIGLGQAKMGFCTDREDINSLCMTVQN 75
      ||| | :|:| | | :|:|:| | | | | | | | | | | | | | | | | | | |
      13 DVGIGALEIYFPQNFVDQNDLEKFNNSGKYTIGLGQQQMFCSDNEDIVISLTYTRK 72

QY      76 LMERNNLSYDCIGRLVEGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGGA 135
      |:|:| | :|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      73 LIETKYISTDSIGCLVGTETIMIDKSKSVKTALMDLF--PGNSDIEGVDIKNACFGGAQA 130

QY      136 VFNANVMI E-----SSSWDGLRGT----- 154
      |:|:| | :|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      131 LHAIDWVTVNHPLDKNAIVVADIAIYEEGPARTCGAGAIAFLICPDASIPIDQFS 190

QY      155 --HMQHAYDFYKP--DMISEYPIVDKLSIQCYLSALDRCTSYVCKKIHAQWQKEGNDKD 210
      ||:|:| | :|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      191 ACHMKNTWDFKXPTPIRSEYFVVDGSLSSYLEAVRMTYTYFISKVNR--HTTGID-- 246

QY      211 FTLNDEGFMI FHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEAFGVDKLEDTYFD 270
      ||| | :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
      247 -GINSFDGVFLHSPFTKVQKGLAVM---NYTDSQLRHQQLNGNV---DHKL DEN--D 296

QY      271 RDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQYSPQQLAG--KRIG 329
      |:|:| | :|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      297 RAQLAKMIELSAQVWKEKTDPLYLVENRRIGNMYTPSLFAQLLAYLA--ADDCVTGEKSIL 354

QY      330 VFSYGSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAPDVAENMKLR 389
      |:|:| | :|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      355 FFAYGSGLASAIFPGRVROT---SNLDKIRQVAIRAIRKLDRIQFTPEEFTEYIQRE 410

QY      390 DTHHLVNIYIPQGSIDSLFEETWYLVRVDEKRRTYARRPTPNDDTLDEGVGLVH 443
      |:|:| | :|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      411 VFLRSKEIPKSPSETSLFPNTYFLDNMDKLYRRSYTLHEEPNG--VQNGNGIHH 462

Db

RESULT 15
HMCS_DICDI STANDARD; PRT; 163 AA.
ID_HMCS_DICDI
AC P54872;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase (EC 2.3.3.10) (HMG-CoA synthase)
DE (3-hydroxy-3-methylglutaryl coenzyme A synthase) (Fragment).
GN HGSA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA Dhar M.S., Hauser L.J., Olins D.E., Olins A.L.;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2004, 05:08:06 ; Search time 762 Seconds
(without alignments)
11161.276 Million cell updates/sec

Title: US-10-622-516-1

Perfect score: 2002

Sequence: 1 cgccctccacgcgactctcgs.....aaaaaaaaaaaaaaaaa 2002

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002s:*
- 7: geneseqn2003as:*
- 8: geneseqn2003bs:*
- 9: geneseqn2003cs:*
- 10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2002	100.0	2002	6	ABSS54409
2	2002	100.0	2002	7	ABX93299
3	1693	84.6	3722	10	ADE76933
4	1605.2	80.2	2540	5	AAS84743
5	1370	68.4	1650	7	ACC62336
6	1370	68.4	1650	7	ACC62328
7	1370	68.4	1650	7	ACC62326
8	1370	68.4	1650	7	ACC62330
9	1370	68.4	1650	7	ACC62338
10	1370	68.4	1650	7	ACC62334
11	1370	68.4	1650	7	ACC62327
12	1370	68.4	1650	7	ACC62332
13	1370	68.4	1650	9	AD338334
14	1322.4	66.1	3008	4	AAH34834
15	1302	65.0	1564	7	ACC62337
16	1302	65.0	1564	7	ACC62339
17	1299.4	64.9	1601	7	ACC62331
18	1299	64.9	1593	7	ACC62329
19	1299	64.9	1612	7	ACC62340
20	1299	64.9	1612	7	ACC62335
21	1295.4	64.7	1608	7	ACC62341
22	1295.4	64.7	1608	7	ACC62333
23	1288	64.3	1685	9	ADD14748

24	1180.6	59.0	3275	8	ACF25333	Acf25333	Rat cytos
25	1180.6	59.0	3275	9	ADB58295	Adb58295	Toxicity-
26	1180.6	59.0	3275	9	ADB52844	Adb52844	Primary r
27	794.6	39.7	1824	2	AAT89089	Aat89089	Avian 3-h
28	476	23.8	28001	6	ABSS54410	Abss54410	Human hyd
29	476	23.8	28001	7	ABX93300	Abx93300	Gene enco
30	460.4	23.0	565	6	AAS61469	Aas61469	Lung smal
31	448.2	22.4	505	7	ABZ84645	Abz84645	Toxicolog
32	436.4	21.8	615	3	AAC79203	Aac79203	Human lun
33	436.4	21.8	615	4	AAD23279	Aad23279	Human lun
34	436.4	21.8	615	9	ADD66553	Add66553	Human lun
35	436.4	21.8	615	9	ADE87807	Ade87807	Human lun
36	423.8	21.2	2093	3	AAC98123	Aac98123	Human col
37	421.8	21.1	2088	4	AAK52002	Aak52002	Human pol
38	421.8	21.1	2377	9	ADE53675	Ades53675	Human pro
39	420.2	21.0	2058	6	ABU69551	Abu69551	Prostate
40	420.2	21.0	2058	6	ABN97315	Abn97315	Gene #381
41	420.2	21.0	2058	9	ACC50182	Acc50182	Breast ca
42	420.2	21.0	2058	9	ADB75349	Adb75349	Prostate
43	420.2	21.0	2132	9	ADE09760	Ade09760	Novel DNA
44	418.2	20.9	1845	8	AAD57512	Aad57512	Human enz
45	413.2	20.6	1994	4	AAH22426	Aah22426	Rat mitoc

ALIGNMENTS

RESULT 1	ABSS4409	ABSS54409	standard; cDNA; 2002 BP.
ID	ABSS4409	ABSS54409	standard; cDNA; 2002 BP.
XX	AC	ABSS4409;	
XX	AC	ABSS4409;	
DT	22-NOV-2002	(first entry)	
XX	XX	Human hydroxymethylglutaryl-coenzyme A synthase protein cDNA.	
XX	XX	Human; gene; ss; chromosome 5; hydroxymethylglutaryl-coenzyme A synthase;	
KW	KW	HMG-CoA synthase; cholesterologenesis; therapeutic; diagnostic; genotype;	
KW	KW	antibody; synthase; carcinoma.	
XX	XX		
OS	OS	Homo sapiens.	
XX	XX		
FH	FH	Key	Location/Qualifiers
FT	FT	5'UTR	1..144
FT	FT	CDS	145..1581
FT	FT		/*tag= b
FT	FT	variation	/product= "Hydroxymethylglutaryl-coenzyme A synthase"
FT	FT		replac(717,T)
FT	FT		/*tag= c
FT	FT	3'UTR	/standard_name= "Single nucleotide polymorphism"
FT	FT		1582..2002
XX	XX		/*tag= d
PN	PN	US6436692-B1.	
XX	XX	20-AUG-2002.	
PD	PD		
XX	XX	29-MAR-2001; 2001US-00819993.	
PF	PF		
XX	XX	29-MAR-2001; 2001US-00819993.	
PR	PR		
XX	XX	(APPL-) APPLERA CORP.	
PA	PA		
XX	XX	Gong F, Yan C, Di Francesco V, Beasley EM;	
PI	PI		
XX	XX	WPI; 2002-689940/74.	
DR	DR	P-PSDB; ABG32726.	
XX	XX		
PT	PT	New isolated nucleic acid molecule encoding hydroxymethylglutaryl-CoA	
PT	PT	synthase, useful as model for the development of human therapeutic	
PT	PT	targets and for identifying therapeutic proteins.	

XX Claim 1; Fig 1; 62pp; English.

CC The invention discloses an isolated nucleic acid molecule encoding a
CC human hydroxymethylglutaryl-coenzyme A (HMG-CoA) synthase which is
CC important in cholesterologenesis. The polynucleotide and polypeptide are
CC useful as models for the development of human therapeutic targets, to aid
CC in the identification of therapeutic proteins and as targets for the
CC development of human therapeutic agents that modulate the activity of the
CC polypeptide in cells and tissues. The polynucleotide is useful for
CC monitoring the effectiveness of modulating compounds on the expression or
CC activity of the enzyme gene in clinical trials and in treatments, in
CC diagnostic assays for qualitative changes in expression of enzyme nucleic
CC acid, to detect mutations in enzyme genes and gene expression products,
CC such a mRNA, for testing an individual for a genotype and as a diagnostic
CC target that can be used to tailor treatment in an individual. The
CC polypeptide is useful to raise antibodies, as a reagent in assays
CC designed to quantitatively determine levels of the protein in biological
CC fluids, as markers for tissues in which the corresponding protein is
CC preferentially expressed (e.g. carcinomas), in biological assays related
CC to members of the synthase subfamily, in drug screening assays, in
CC competition binding assays, in cell-based or cell-free systems, in
CC pharmacogenomic analysis and for treating a disorder characterised by an
CC absence of, inappropriate, or unwanted expression of the polypeptide. The
CC sequence presented is the human HMG-CoA synthase protein cDNA, the gene
CC for which is located on chromosome 5

XX Sequence 2002 BP; 590 A; 380 C; 472 G; 560 T; 0 U; 0 Other;

Query Match 100.0%; Score 2002; DB 6; Length 2002;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCCTCCCAAGCAGCTCTTCGGCAGTAGTCCCGAGTCGGGTGGCGGCTATAAAGCTGT 60
DB 1 CGCCTCCCAAGCAGCTCTTCGGCAGTAGTCCCGAGTCGGGTGGCGGCTATAAAGCTGT 60
QY 61 AGCGAAGGGGAGGCGCGCGGAGCTGTCTTCGTGGCTCACTCCCTTCTCTGTCGCCG 120
DB 61 AGCGAAGGGGAGGCGCGCGGAGCTGTCTTCGTGGCTCACTCCCTTCTCTGTCGCCG 120
QY 121 CTGGTCAAGCTTGTCTTCAACCATGCTGGATCACTTCTTGAATGAGAAGCTTGC 180
DB 121 CTGGTCAAGCTTGTCTTCAACCATGCTGGATCACTTCTTGAATGAGAAGCTTGC 180
QY 181 TGGCCAAAAGATGTGGGAAATTGTTGCCCTTGAAGATCTATTTCTCTCAATATGTGAT 240
DB 181 TGGCCAAAAGATGTGGGAAATTGTTGCCCTTGAAGATCTATTTCTCTCAATATGTGAT 240
QY 241 CAAGCAGAGTTGGAAAAATATGATGTGTAGATGTGAAAGTATACCATGGCTTGGGC 300
DB 241 CAAGCAGAGTTGGAAAAATATGATGTGTAGATGTGAAAGTATACCATGGCTTGGGC 300
QY 301 CAGGCCAAGATGGGCTTCTGCACAGATAGAGAATATTACTCTTTGCATGACTGTG 360
DB 301 CAGGCCAAGATGGGCTTCTGCACAGATAGAGAATATTACTCTTTGCATGACTGTG 360
QY 361 GTTCAGAACTTTATGAGAGAAATAACCTTTCTATGATTGCAATTGGCGGCTGGAAGTT 420
DB 361 GTTCAGAACTTTATGAGAGAAATAACCTTTCTATGATTGCAATTGGCGGCTGGAAGTT 420
QY 421 GGAACAGAGCAATCATCGACAATCAAAAGTCTGTGAAGACTAATTGATGCACTGTTT 480
DB 421 GGAACAGAGCAATCATCGACAATCAAAAGTCTGTGAAGACTAATTGATGCACTGTTT 480
QY 481 GAAGAGTCTGGGAATACAGATATAGAAGAAATCGACACAACCTAATGCATGTATGGAGC 540
DB 481 GAAGAGTCTGGGAATACAGATATAGAAGAAATCGACACAACCTAATGCATGTATGGAGC 540
QY 541 ACAGCTGCTGCTTCAATGCTGTAACTGGATGAGTCAGCTCTTGGAGTGGGCTTGGT 600
DB 541 ACAGCTGCTGCTTCAATGCTGTAACTGGATGAGTCAGCTCTTGGAGTGGGCTTGGT 600

QY 601 GGGACACATATGCAACATGCGCTATGATTTTACAAAGCTGATATGCTATCTGAATATCCT 660
DB 601 GGGACACATATGCAACATGCGCTATGATTTTACAAAGCTGATATGCTATCTGAATATCCT 660
QY 661 ATAGTAGATGAAAACTCTCCATACAGTGTCACTCACTGATTAAGCCGCTGCTACTCT 720
DB 661 ATAGTAGATGAAAACTCTCCATACAGTGTCACTCACTGATTAAGCCGCTGCTACTCT 720
QY 721 GTTAAGTCAAAAAAGATCCATGCCCCAGTGGCAGAAAGAGGAAATGATTAAGATTTTACC 780
DB 721 GTTAAGTCAAAAAAGATCCATGCCCCAGTGGCAGAAAGAGGAAATGATTAAGATTTTACC 780
QY 781 TTGAATGATTTTGGCTTCATGATCTTCACTCACCATATTTGTAAGTGTTCAGAAATCT 840
DB 781 TTGAATGATTTTGGCTTCATGATCTTCACTCACCATATTTGTAAGTGTTCAGAAATCT 840
QY 841 CTAGCTCGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAGATAAATAAGTATC 900
DB 841 CTAGCTCGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAGATAAATAAGTATC 900
QY 901 TATAGTGGCTTGAAGCCTTTGGGATGTTAAATTGAAAGACACCTACTTTGATAGAGAT 960
DB 901 TATAGTGGCTTGAAGCCTTTGGGATGTTAAATTGAAAGACACCTACTTTGATAGAGAT 960
QY 961 GTGGAAGAGGCAATTTAAGAAGCTAGCTCTGAACCTCTCACTCAGAAAACAAGGCATCT 1020
DB 961 GTGGAAGAGGCAATTTAAGAAGCTAGCTCTGAACCTCTCACTCAGAAAACAAGGCATCT 1020
QY 1021 TTACTGTATCAAAATCAAAATGAAATATGTACACATCTTCAATATATGTTCCCTTGCA 1080
DB 1021 TTACTGTATCAAAATCAAAATGAAATATGTACACATCTTCAATATATGTTCCCTTGCA 1080
QY 1081 TCTGTTTACACACAGTACTCACTCAGCAATTAGCAGGAAAGAAATGGAGTGTCTTCT 1140
DB 1081 TCTGTTTACACACAGTACTCACTCAGCAATTAGCAGGAAAGAAATGGAGTGTCTTCT 1140
QY 1141 TATGTTCTGCTTGGCTGCCACTCTGTACTCTCTTAAGTCAACAAAGTCTACACCCG 1200
DB 1141 TATGTTCTGCTTGGCTGCCACTCTGTACTCTCTTAAGTCAACAAAGTCTACACCCG 1200
QY 1201 GGGTCTGCTCTTGATTAATAATAACAGCAAGTTATGTATCTTAAATCAAGGCTTGATTCA 1260
DB 1201 GGGTCTGCTCTTGATTAATAATAACAGCAAGTTATGTATCTTAAATCAAGGCTTGATTCA 1260
QY 1261 AGAAGTGGTGGCACCAAGATGCTTGGCTGAAAAACATGAAGCTCAGAGAGACACCCAT 1320
DB 1261 AGAAGTGGTGGCACCAAGATGCTTGGCTGAAAAACATGAAGCTCAGAGAGACACCCAT 1320
QY 1321 CATTTGTCAACTATATTTCCCAAGGTTCAATAGATTCACTTTGAAGAAAGTGTATC 1380
DB 1321 CATTTGTCAACTATATTTCCCAAGGTTCAATAGATTCACTTTGAAGAAAGTGTATC 1380
QY 1381 TTAGTTAGGTTGATGAAAAAGCAGAAAGAACTTACGCTGGCGTCCCAATATGAT 1440
DB 1381 TTAGTTAGGTTGATGAAAAAGCAGAAAGAACTTACGCTGGCGTCCCAATATGAT 1440
QY 1441 GACACTTTGATGAAGGATGAGACTTGTGCAATTCAAACATAGCAACTGAGCATATTCCA 1500
DB 1441 GACACTTTGATGAAGGATGAGACTTGTGCAATTCAAACATAGCAACTGAGCATATTCCA 1500
QY 1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGAGAACTGAAAGCACTGTC 1560
DB 1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGAGAACTGAAAGCACTGTC 1560
QY 1561 ATTAGTAATGGGAAACATTAAGATACTCTGTGAGGTGCAAGACTTCAAGGTGGGTGGC 1620
DB 1561 ATTAGTAATGGGAAACATTAAGATACTCTGTGAGGTGCAAGACTTCAAGGTGGGTGGC 1620
QY 1621 ATGGGTGGGGTATGGGAAGTGAAGGAAATGGGATATCTGGGATTAATTTTAAAGGA 1680
DB 1621 ATGGGTGGGGTATGGGAAGTGAAGGAAATGGGATATCTGGGATTAATTTTAAAGGA 1680
QY 1681 TTACATGTTATGTAAATTTTATGTGACTGACATGGAAGCCTGATGACTATCTGTACTT 1740

|||||
Db 1681 TTAACATGTATGTAATTTTATGTGACTGACATGAGCCCTGGATGACTATCGTACTT 1740
QY 1741 GGGAAAGTCTCTTGTCTATTTGCTGACATGCTTCCGTGTGTGCTGCGCAATGCCA 1800
Db 1741 GGGAAAGTCTCTTGTCTATTTGCTGACATGCTTCCGTGTGTGCTGCGCAATGCCA 1800
QY 1801 ATGTACTCGAATGATGTTAAGGCTCTGTAAACCTTCATACCTTTGGCCATTGTATG 1860
Db 1801 ATGTACTCGAATGATGTTAAGGCTCTGTAAACCTTCATACCTTTGGCCATTGTATG 1860
QY 1861 CATGATGTTGTTGTTTAAACAATGATTAATGTAATGTTGTTACTTCTGTCAGAGAAGCA 1920
Db 1861 CATGATGTTGTTGTTTAAACAATGATTAATGTAATGTTGTTACTTCTGTCAGAGAAGCA 1920
QY 1921 GAGGTACTAATCTCCAATTAATAATTTTAAACATGTAAATAAAAAAAAAAAAAA 1980
Db 1921 GAGGTACTAATCTCCAATTAATAATTTTAAACATGTAAATAAAAAAAAAAAAAA 1980
QY 1981 AAAAAAAAAAAAAAAAAAAAAA 2002
Db 1981 AAAAAAAAAAAAAAAAAAAAAA 2002

RESULT 2
ABX93299
ID ABX93299 standard; cDNA; 2002 BP.
XX
AC ABX93299;
DT 05-JUN-2003 (first entry)
XX
DE cDNA encoding human HMG-CoA synthase-like enzyme.
XX
KW Human; cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase;
KW HMG-CoA synthase; immune response; drug screening assay;
KW pharmacogenomic analysis; chromosome 5; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..144
FT /*tag= a
FT 145..1581
FT CDS /*tag= b
FT /product= "HMG-CoA synthase-like enzyme"
FT 3'UTR 1582..2002
FT /*tag= c
XX
PN US2002173018-A1.
XX
PD 21-NOV-2002.
XX
PF 12-JUL-2002; 2002US-00193295.
XX
PR 29-MAR-2001; 2001US-00819993.
XX
PA (APPL-) APPLERA CORP.
XX
PI Gong F, Yan C, Di Francesco V, Beasley EM;
XX
DR MPI; 2003-352594/33.
DR P-PSDB; ABU08379.
XX
PT New isolated human synthase peptide and gene encoding the enzyme, useful
PT as models for developing human therapeutic targets, aid in the
PT identification of therapeutic proteins and as therapeutic targets.
XX
PS Claim 22; Fig 1A; 65pp; English.
XX
CC The present invention relates to the isolation of a novel human enzyme
CC that shows a high degree of similarity to human cytoplasmic 3-hydroxy-3-
CC methylglutaryl coenzyme A (HMG-CoA) synthase. The invention also

CC discloses polynucleotide sequences encoding the novel enzyme of the
CC invention. Both the polypeptide and polynucleotide sequences are useful
CC as models for the development of human therapeutics, for identifying
CC therapeutic proteins, as targets for development of human therapeutic
CC agents, and as query sequences to perform a search against sequence data
CC bases to identify other family members of related sequences. The
CC polypeptide is useful to raise antibodies or to elicit another immune
CC response, as a reagent in assays designed to quantitatively determine
CC levels of the protein in biological fluids, as markers for tissues in
CC which the corresponding protein is preferentially expressed, in drug
CC screening assays, in cell-based or cell-free systems, to identify
CC compounds that modulate synthase activity of the protein in its natural
CC state, or an altered form that causes the specific disease or pathology
CC associated with the synthase, to screen a compound for the ability to
CC stimulate or inhibit interaction between the synthase protein and a
CC molecule that normally interacts with the synthase protein, and in
CC pharmacogenomic analysis. The polynucleotide is useful for monitoring the
CC effectiveness of modulating compounds on the expression or activity of
CC the human synthase gene in clinical trials or in a treatment regimen, in
CC diagnostic assays for qualitative changes in a human synthase nucleic
CC acid that leads to a pathology, for testing an individual for a genotype
CC that while not necessarily causing a disease, nevertheless affects the
CC treatment modality, and as antisense constructs to control human synthase
CC gene expression in cells, tissues and organisms. The present sequence
CC encodes human HMG-CoA synthase-like enzyme. The gene encoding the enzyme
CC is located on chromosome 5
XX
SQ Sequence 2002 BP; 590 A; 380 C; 472 G; 560 T; 0 U; 0 Other;

Query Match 100.0%; Score 2002; DB 7; Length 2002;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCTCCAGGAGTCTCGGAGTGCAGGAGTCCGGGTGGGCTATAAAGCTGT 60
Db 1 CGCTCCAGGAGTCTCGGAGTGCAGGAGTCCGGGTGGGCTATAAAGCTGT 60
QY 61 AGCGAAGGGAGGCGCGGAGTCTCTTCTGCTCACTCCCTTCTCTGCTGCC 120
Db 61 AGCGAAGGGAGGCGCGGAGTCTCTTCTGCTCACTCCCTTCTCTGCTGCC 120
QY 121 CTGGTCACGCTTGCTCTTCAACCATGCTGATCATTCTTGAATGAGAAGCTTGC 180
Db 121 CTGGTCACGCTTGCTCTTCAACCATGCTGATCATTCTTGAATGAGAAGCTTGC 180
QY 181 TGGCCAAAAGATGTGGAAATGTGCCCTTGAGATCTAATTTCTCTCAATATGTTGAT 240
Db 181 TGGCCAAAAGATGTGGAAATGTGCCCTTGAGATCTAATTTCTCTCAATATGTTGAT 240
QY 241 CAAGCAGAGTTGAAAAATATGATGTTAGATGCTGGAAGTATACCATTTGGCTGGGC 300
Db 241 CAAGCAGAGTTGAAAAATATGATGTTAGATGCTGGAAGTATACCATTTGGCTGGGC 300
QY 301 CAGGCCAAGATGGGCTTCTGCACAGATAGAGAAGATATTAATCTCTTGCATGACTGTG 360
Db 301 CAGGCCAAGATGGGCTTCTGCACAGATAGAGAAGATATTAATCTCTTGCATGACTGTG 360
QY 361 GTTCAGAACTTATGGAGAGAAATAACCTTCTCATGATTGCAATGGCGGCTGGAAGTT 420
Db 361 GTTCAGAACTTATGGAGAGAAATAACCTTCTCATGATTGCAATGGCGGCTGGAAGTT 420
QY 421 GGAACAGAGCAATCATCGACAATCAAAAGTCTGTGAAGACTAATTTGATGAGCTGTTT 480
Db 421 GGAACAGAGCAATCATCGACAATCAAAAGTCTGTGAAGACTAATTTGATGAGCTGTTT 480
QY 481 GAAGAGTCTGGGATACAGATATAGAAAGAAATCGACACAACTAATGCTATGAGAGC 540
Db 481 GAAGAGTCTGGGATACAGATATAGAAAGAAATCGACACAACTAATGCTATGAGAGC 540
QY 541 ACAGCTGCTGCTTCAATGCTGTTAACTGATTGAGTCCAGCTCTTGGGATGGGCTTGT 600
Db 541 ACAGCTGCTGCTTCAATGCTGTTAACTGATTGAGTCCAGCTCTTGGGATGGGCTTGT 600

QY 601 GGGACACATATGCAACATGCGCTATGATTTTACAAGCGCTGATATGCTATCTGAATATCCT 660
Db 601 GGGACACATATGCAACATGCGCTATGATTTTACAAGCGCTGATATGCTATCTGAATATCCT 660
QY 661 ATAGTAGATGAAAACTCTCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTACTCT 720
Db 661 ATAGTAGATGAAAACTCTCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTACTCT 720
QY 721 GTCCTACTGCAAAAAAGATCCATGCCAGTGGCAGAAAAGAGGAAATGATATAAGATTTTACC 780
Db 721 GTCCTACTGCAAAAAAGATCCATGCCAGTGGCAGAAAAGAGGAAATGATATAAGATTTTACC 780
QY 781 TTGAATGATTTTGGCTTCATGATCTTTCACTCACCATAATTGTAAACTGGTTCAGAAATCT 840
Db 781 TTGAATGATTTTGGCTTCATGATCTTTCACTCACCATAATTGTAAACTGGTTCAGAAATCT 840
QY 841 CTAGCTGGGATGTTGCTGATGACTTCTTAATGACCAGAAATAGAGATAAAATAGTATC 900
Db 841 CTAGCTGGGATGTTGCTGATGACTTCTTAATGACCAGAAATAGAGATAAAATAGTATC 900
QY 901 TATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGAT 960
Db 901 TATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGAT 960
QY 961 GTGAGAAGGCATTTATGAAGGCTAGCTCTGAACCTCTTCACTCAGAACAGAAAACAAGGCATCT 1020
Db 961 GTGAGAAGGCATTTATGAAGGCTAGCTCTGAACCTCTTCACTCAGAACAGAAAACAAGGCATCT 1020
QY 1021 TTACTTGATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGGTTCCCTTGCA 1080
Db 1021 TTACTTGATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGGTTCCCTTGCA 1080
QY 1081 TCTGTTCTTAGCACAGTACTCACTCAGCAATTTAGCAGGGAAGAAATTGGAGTGTCTTCT 1140
Db 1081 TCTGTTCTTAGCACAGTACTCACTCAGCAATTTAGCAGGGAAGAAATTGGAGTGTCTTCT 1140
QY 1141 TATGTTCTGGTTGGCTGCCACTCTGTACTCTCTTAAGTCAACACAGAATGCTACACCG 1200
Db 1141 TATGTTCTGGTTGGCTGCCACTCTGTACTCTCTTAAGTCAACACAGAATGCTACACCG 1200
QY 1201 GGGTCTGCTCTTGATAAAATAACAGCAAGTTATGTGATCTTAAATCAAGCCTTGATTCA 1260
Db 1201 GGGTCTGCTCTTGATAAAATAACAGCAAGTTATGTGATCTTAAATCAAGCCTTGATTCA 1260
QY 1261 AGAAGTGTGTGGCACCAAGATGTCTTCCGTGAATAACATGAAGCTCAGAGAGACACCCAT 1320
Db 1261 AGAAGTGTGTGGCACCAAGATGTCTTCCGTGAATAACATGAAGCTCAGAGAGACACCCAT 1320
QY 1321 CATTGGTCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTAC 1380
Db 1321 CATTGGTCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTAC 1380
QY 1381 TTAGTTAGGGTGATGAAAAAGCACAGAAAGAACTTACGCTCGGCGTCCCACTCCAAATGAT 1440
Db 1381 TTAGTTAGGGTGATGAAAAAGCACAGAAAGAACTTACGCTCGGCGTCCCACTCCAAATGAT 1440
QY 1441 GACACTTTGGATGAAGAGTAGGAACTTGTGCATTCAAAACATAGCAACTGAGCATATTCCA 1500
Db 1441 GACACTTTGGATGAAGAGTAGGAACTTGTGCATTCAAAACATAGCAACTGAGCATATTCCA 1500
QY 1501 AGCCCTGCCAAGAAAGTACCAAGAACTCCCTGCCACAGCAGAACTGAGAGCACTGTC 1560
Db 1501 AGCCCTGCCAAGAAAGTACCAAGAACTCCCTGCCACAGCAGAACTGAGAGCACTGTC 1560
QY 1561 ATTAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGC 1620
Db 1561 ATTAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGC 1620
QY 1621 ATGGGGTGGGGTATGGAAACAGTTGGAGAGATGGGATATCTGGGGATAATTTTAAAGGA 1680
Db 1621 ATGGGGTGGGGTATGGAAACAGTTGGAGAGATGGGATATCTGGGGATAATTTTAAAGGA 1680
QY 1681 TTACATGTATGTAAATTTTATGTGACTGACATGAGCGCTGGATGACTATCGTACTT 1740
Db 1681 TTACATGTATGTAAATTTTATGTGACTGACATGAGCGCTGGATGACTATCGTACTT 1740

Db 1681 TTACATGTATGTAAATTTTATGTGACTGACATGAGCGCTGGATGACTATCGTACTT 1740
QY 1741 GGGAAAGTCTCTTTGCTCTATTTTGCTGACATGCTTCCTGTGTGTGGCCAATGCCAA 1800
Db 1741 GGGAAAGTCTCTTTGCTCTATTTTGCTGACATGCTTCCTGTGTGTGGCCAATGCCAA 1800
QY 1801 ATGTACTGGAATGATGTTAAGGGCTCTGTAAACCTTCATACCTCTTTGGCCAATTTGTATG 1860
Db 1801 ATGTACTGGAATGATGTTAAGGGCTCTGTAAACCTTCATACCTCTTTGGCCAATTTGTATG 1860
QY 1861 CATGATGTTTGGTTTAAACATGGTATATAAGATTGTGTACTTCTGTCAAGAAAGCA 1920
Db 1861 CATGATGTTTGGTTTAAACATGGTATATAAGATTGTGTACTTCTGTCAAGAAAGCA 1920
QY 1921 GAGGTACTAATCTCCAAATTAATAATTTTAAACATGTATAAAAAA 1980
Db 1921 GAGGTACTAATCTCCAAATTAATAATTTTAAACATGTATAAAAAA 1980
QY 1981 AAAAAAAAAAAAAAAAAA 2002
Db 1981 AAAAAAAAAAAAAAAAAA 2002

RESULT 3
ADE76933
ID ADE76933 standard; cDNA; 3722 BP.
XX
AC ADE76933;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human cDNA differentially expressed in a liver disorder #70.
XX
KW human; es; gene; liver disorder; hyperlipidaemia; hypertension;
KW type II diabetes; tumour; liver; inflammatory disorder;
KW immune response disorder; high-throughput screening;
KW differential gene expression; gene therapy.
XX
OS Homo sapiens.
XX
PN US2003108871-A1.
XX
PD 12-JUN-2003.
XX
PF 30-JUL-2001; 2001US-00919039.
XX
PR 28-JUL-2000; 2000US-0222113P.
XX
PA (KASE/) KASER M R.
XX
PI Kaser MR;
XX
XX WPI; 2004-031227/03.
XX
PT Composition comprising several cDNAs that are differentially expressed in
PT treated human C3A liver cell cultures, useful for treating liver
PT disorders.
XX
PS Claim 1; SEQ ID NO 98; 41bp; English.
XX
CC The invention relates to a composition comprising several cDNAs that are
CC differentially expressed in a liver disorder. The composition is useful
CC for treating liver disorder such as hyperlipidaemia, hypertension, type
CC II diabetes, tumours of the liver and disorders of the inflammatory and
CC immune response. The composition is useful for a high-throughput method
CC of screening several molecules or compounds to identify a ligand which
CC specifically binds a cDNA. A protein encoded by the cDNA is useful for a
CC high-throughput method for using a protein to screen several molecules or
CC compounds to identify at least one ligand which specifically binds the
CC protein which involves combining the protein encoded by the cDNA with
CC several of molecules or compounds under conditions to allow specific
CC binding, and detecting specific binding between the protein and a


```

      |||
Db      2218 ACAGAAA 2224

RESULT 4
AAS84743
ID      AAS84743 standard; cDNA; 2540 BP.
XX
AC      AAS84743;
XX
DT      13-FEB-2002 (first entry)
XX
DE      DNA encoding novel human diagnostic protein #20547.
XX
KW      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM      food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS      Homo sapiens.
XX
PN      WO200175067-A2.
XX
PD      11-OCT-2001.
XX
PF      30-MAR-2001; 2001WO-US008631.
XX
PR      31-MAR-2000; 2000US-00540217.
PR      23-AUG-2000; 2000US-00649167.
XX
PA      (HYSE-) HYSEQ INC.
XX
PI      Drmanac RT, Liu C, Tang YT;
XX
DR      WPI; 2001-639362/73.
DR      P-PSDB; ABG20556.
XX
PT      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity.
XX
PS      Claim 1; SEQ ID NO 20547; 103bp; English.
XX
CC      The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC      sequences. (I) is useful as hybridisation probes, polymerase chain
CC      reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC      and in recombinant production of (II). The polynucleotides are also used
CC      in diagnostics as expressed sequence tags for identifying expressed
CC      genes. (I) is useful in gene therapy techniques to restore normal
CC      activity of (II) or to treat disease states involving (II). (II) is
CC      useful for generating antibodies against it, detecting or quantitating a
CC      polypeptide in tissue, as molecular weight markers and as a food
CC      supplement. (II) and its binding partners are useful in medical imaging
CC      of sites expressing (II). (I) and (II) are useful for treating disorders
CC      involving aberrant protein expression or biological activity. The
CC      polypeptide and polynucleotide sequences have applications in
CC      diagnostics, forensics, gene mapping, identification of mutations
CC      responsible for genetic disorders or other traits to assess biodiversity
CC      and to produce other types of data and products dependent on DNA and
CC      amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC      coding sequences of the invention. Note: The sequence data for this
CC      patent did not appear in the printed specification, but was obtained in
CC      electronic format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 2540 BP; 732 A; 474 C; 621 G; 713 T; 0 U; 0 Other;

Query Match      80.2%; Score 1605.2; DB 5; Length 2540;
Best Local Similarity 89.6%; Pred. NO. 3.3e-305;
Matches 1921; Conservative 0; Mismatches 18; Indels 205; Gaps 10;

QY      25 GCCGGAGTCGGGTGGTTGGCGGCTATAAAGCTGTAGCGAAGGGAGGCGCGGCGACT 84
      |||
Db      5 GTCGGAGTCGGGTGGTTGGCGGCTATAAAGCTGTGTGGCGAAGGGAGGCGCGGCGACT 64
```

```

QY      85 GTCCTTTCGTGGCTCACTCCCTTTCCTGTGTCGCCGCTCGGTCAACGCTT----- 133
      |||
Db      65 GTCCTTTCGTGGCTCACTCCCTTTCCTGTGTCGCCGCTCGGTCAACGCTTGTGCCGAAGG 124
      |||
QY      134 -----GCTCTTTCACCA 145
      |||
Db      125 AGGAAACAGTGCACAGACCTGGAGACTGCAGTTCTCTATCCTTCACACAGCTCTTTCACCA 184
      |||
QY      146 TGCCCTGATCACTTCCCTTGAATGCAGAAAGCTTGCTGGCCAAAGATGTGGGAATTGTTG 205
      |||
Db      185 TGCCCTGATCACTTCCCTTGAATGCAGAAAGCTTGCTGGCCAAAGATGTGGGAATTGTTG 244
      |||
QY      206 CCCTTGAGATCTAATTTTCTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGATG 265
      |||
Db      245 CCCTTGAGATCTAATTTTCTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGATG 304
      |||
QY      266 GTGTAGATGCTGGAAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGGGCTTTCGACAG 325
      |||
Db      305 GTGTAGATGCTGGAAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGGGCTTTCGACAA 364
      |||
QY      326 ATAGAGAAGATATTAATCTCTTTCATGATGACTGTGTTCAAGATCTTATGGAGAAATA 385
      |||
Db      365 ATAGAGAAGATATTAATCTCTTTCATGATGACTGTGTTCAAGATCTTATGGAGAAATA 424
      |||
QY      386 ACCTTTCATGATTTGCATTTGGGCGGCTGGAAAGTTGGAAACAGAGCAATCATGCACAAAT 445
      |||
Db      425 ACCTTTCATGATTTGCATTTGGGCGGCTGGAAAGTTGGAAACAGAGCAATCATGCACAAAT 484
      |||
QY      446 CAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATAG 505
      |||
Db      485 CAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATAG 544
      |||
QY      506 AAGGAATGCACACAACATTAATGCATGCTATGAGGACACAGCTGCTGCTTCAAATGCTGTTA 565
      |||
Db      545 AAGGAATGCACACAACATTAATGCATGCTATGAGGACACAGCTGCTGCTTCAAATGCTGTTA 604
      |||
QY      566 ACTGATTGAGTCCAGCTCTTGGGATGGGC----- 595
      |||
Db      605 ACTGATTGAGTCCAGCTCTTGGGATGGACGATATGCCCTGGGTAGTTGCAGAGATATT 664
      |||
QY      596 ----- 595
      |||
Db      665 GCTGTATATGCCACAGAAATGCTAGCCTACAGGTGGAGTTGGACAGTAGCTCTTGC 724
      |||
QY      596 -----TTCGTGGACACA 608
      |||
Db      725 TAAATGGGGCCAAATAGCTCCTTTAATTTTGAACGAGGGGTTTCGTGGGACACA 784
      |||
QY      609 TATGCAACATGCCATATGATTTTACAAGCCGATATGCTAATCTGAATATCCATATAGTAGA 668
      |||
Db      785 TATGCAACATGCCATATGATTTTACAAGCCGATATGCTAATCTGAATATCCATATAGTAGA 844
      |||
QY      669 TGGAAAACTCTCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTACTGTCTACTG 728
      |||
Db      845 TGGAAAACTCTCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTACTGTCTACTG 904
      |||
QY      729 CAAAAAGATCCATGCCCCAGTGGCAGAAAGGGAATGATAAAGATTTTACCTTGAATGA 788
      |||
Db      905 CAAAAAGATCCATGCCCCAGTGGCAGAAAGGGAATGATAAAGATTTTACCTTGAATGA 964
      |||
QY      789 TTTTGGCTTCATGATCTTTCACTCACCATATTTGTAAACTGGTTCAGAAATCTCTAGCTCG 848
      |||
Db      965 TTTTGGCTTCATGATCTTTCACTCACCATATTTGTAAACTGGTTCAGAAATCTCTAGCTCG 1024
      |||
QY      849 GATGTTGCTGAATGACTTCTTAAATGACCAAGATAGAGATAAATAGTATCTATAGTGG 908
      |||
Db      1025 GATGTTGCTGAATGACTTCTTAAATGACCAAGATAGAGATAAATAGTATCTATAGTGG 1084
      |||
QY      909 CCTGGAAGCCTTTGGGAGTGTAAATTAGAAACAACCTACTTTGATAGAGATGTGAGAA 968
      |||
Db      1085 CCTGGAAGCCTTTGGGAGTGTAAATTAGAAACAACCTACTTTGATAGAGATGTGAGAA 1144
```

QY	969	GGCATTATGAA	GCGTAGCT	CTGAACTCTT	CAGTTCAGAAAA	CAAAAGCATCTT	TACTGT	1028
Db	1145	GGCATTATGAA	GCGTAGCT	CTGAACTCTT	CAGTTCAGAAAA	CAAAAGCATCTT	TACTGT	1204
QY	1029	ATCAAAATCAAAA	TGAAATATGT	TACACATCTT	CAGTATATGTT	CCCTTGCACTGTCT		1088
Db	1205	ATCAAAATCAAAA	TGAAATATGT	TACACATCTT	CAGTATATGTT	CCCTTGCACTGTCT		1264
QY	1089	AGCACAGTACTCA	CTCAGCAATT	TGACAGGAA	GAAATTTGAGT	GTCTTTCTTA	TGTTTC	1148
Db	1265	AGCACAGTACTCA	CTCAGCAATT	TGACAGGAA	GAAATTTGAGT	GTCTTTCTTA	TGTTTC	1324
QY	1149	T-GGTTGGCTG	CCACTCTGT	ACTCTTTAA	AGTCACA	CAAGATGCT	ACACGGGGCTG	1207
Db	1325	TGGGTTGGCTG	CCACTCTGT	ACTCTTTAA	AGTCACA	CAAGATGCT	ACACGGGGCTG	1384
QY	1208	CTCTTGAT-AAA	ATACAGCAAGT	TATGTGATCTT	TAATCAAGGCTT	GATTTCAAGACT		1266
Db	1385	CTCTTGAT-AAA	ATACAGCAAGT	TATGTGATCTT	TAATCAAGGCTT	GATTTCAAGACT		1444
QY	1267	GGTGTGGCACC	AGATGTCTT	CGCTGAAA	ACATGAAGCTC	AGAGAGACACCC	ATCATTTG	1326
Db	1445	GGTGTGGCACC	AGATGTCTT	CGCTGAAA	ACATGAAGCTC	AGAGAGACACCC	ATCATTTG	1504
QY	1327	GTCACATATAT	TCCCCAGGGT	TCATAGATTCA	CTTTGAA	GGAACGTGTACTTA	GT	1386
Db	1505	GTCACATATAT	TCCCCAGGGT	TCATAGATTCA	CTTTGAA	GGAACGTGTACTTA	GT	1564
QY	1387	AGGGTGATGAAA	AGCAGAAAGAAC	-TTACGCTCGGCGT	CCCACTCCAA	ATGATGACAC		1445
Db	1565	AGGGTGATGAAA	AGCAGAAAGAAC	-TTACGCTCGGCGT	CCCACTCCAA	ATGATGACAC		1624
QY	1446	TTTGATGAAG	AGTAGGACTT	TGCAATTC	AAACATAGCA	CTGAGCATAT	TCCAAGCCC	1505
Db	1625	TTTGATGAAG	AGTAGGACTT	TGCAATTC	AAACATAGCA	CTGAGCATAT	TCCAAGCCC	1684
QY	1506	TGCCAAGAA	AGTACCAAGACT	CCCCCTGCC	ACACAGCAGAA	CCTGAAGCAGCTGT	CAATTAG	1565
Db	1685	TGCCAAGAA	AGTACCAAGACT	CCCCCTGCC	ACACAGCAGAA	CCTGAAGCAGCTGT	CAATTAG	1744
QY	1566	TAATGGGGA	CACTTAAGATA	CTCTGTAGG	TGCAAGACTT	CAAGGCTGGGGTGG	GCATGGG	1625
Db	1745	TAATGGGGA	CACTTAAGATA	CTCTGTAGG	TGCAAGACTT	CAAGGCTGGGGTGG	GCATGGG	1804
QY	1626	GTGGGGTAT	TGGAA	CAGTTGGA	GGAATGGA	TATCTGGGATTA	TTTAAAGATTACA	1685
Db	1805	GTGGGGTAT	TGGAA	CAGTTGGA	GGAATGGA	TATCTGGGATTA	TTTAAAGATTACA	1864
QY	1686	TGTTATGTA	AAATTTTAA	TGTGACTG	ACAT-GGAGCCTG	GA	CTATCGTACTTGGGA	1744
Db	1865	TGTTATGTA	AAATTTTAA	TGTGACTG	ACAT-GGAGCCTG	GA	CTATCGTACTTGGGA	1924
QY	1745	AAGTCTCTT	GCTCTATT	TGCTGACATG	-CTTCCGTGTG	TGTGCTGGCCAATGCC	-AAA	1801
Db	1925	AAGTCTCTT	GCTCTATT	TGCTGACATG	-CTTCCGTGTG	TGTGCTGGCCAATGCC	-AAA	1984
QY	1802	TGTA	CTCGAATGATG	TAAAGGCTCTG	TAAACCTTCAT	ACCTCTTGGCCAT	TTGTATGC	1861
Db	1985	GTA	CTCGAATGATG	TAAAGGCTCTG	TAAACCTTCAT	ACCTCTTGGCCAT	TTGTATGC	2044
QY	1862	ATGATGTTT	GGTTTTTAA	CATG	GTATAT-GAATTGT	GTA	CTTCTGTCA	GAA-1919
Db	2045	ATGATGTTT	GGTTTTTAA	CATG	GTATAT-GAATTGT	GTA	CTTCTGTCA	GAA-2104
QY	1920	AGAGGTA	CTAATCTCCA	ATTAAAAATTTT	TTA	CATGTAAAA		1963
Db	2105	AGAGGTA	CTAATCTCCA	ATTAAAAATTTT	TTA	CATGTAAAA		2148

RESULT 5
ACC62336
ID ACC62336 standard; cDNA; 1650 BP.
XX

AC	ACC62336;	
XX		
DT	23-JUN-2003	(first entry)
XX		
DE	Human NOV41k encoding cDNA SEQ ID NO:201.	
XX		
KW	Human; NOVX; antiatherosclerotic; hypotensive; cardiac; dermatological;	
KW	Human; immunosuppressive; cytostatic; antidiabetic; antifertility;	
KW	haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;	
KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;	
KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma	
KW	congenital heart defect; aortic stenosis; valve disease; transplantation	
KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;	
KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;	
KW	fertility; haemophilia; hypercoagulation; graft versus host disease;	
KW	idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;	
KW	Crohn's disease; multiple sclerosis; infectious disease; cancer;	
KW	cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;	
KW	immune disorder; haematopoietic disorder; dyslipidaemia;	
KW	metabolic syndrome X; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO2003023001-A2.	
XX		
PD	20-MAR-2003.	
XX		
PF	09-SEP-2002; 2002WO-US028538.	
PR		
XX		
PR	07-SEP-2001; 2001US-0318120P.	
PR	07-SEP-2001; 2001US-0318184P.	
PR	10-SEP-2001; 2001US-0318430P.	
PR	17-SEP-2001; 2001US-0322636P.	
PR	17-SEP-2001; 2001US-0322781P.	
PR	17-SEP-2001; 2001US-0322816P.	
PR	17-SEP-2001; 2001US-0322817P.	
PR	19-SEP-2001; 2001US-0323519P.	
PR	20-SEP-2001; 2001US-0323631P.	
PR	20-SEP-2001; 2001US-0323636P.	
PR	25-SEP-2001; 2001US-0324969P.	
PR	25-SEP-2001; 2001US-0325091P.	
PR	26-SEP-2001; 2001US-0324990P.	
PR	14-DEC-2001; 2001US-0341144P.	
PR	26-FEB-2002; 2002US-0359599P.	
PR	05-MAR-2002; 2002US-0361663P.	
PR	03-MAY-2002; 2002US-0377908P.	
PR	17-MAY-2002; 2002US-0381483P.	
PR	29-MAY-2002; 2002US-0383863P.	
PR	02-JUL-2002; 2002US-0393332P.	
PR	17-JUL-2002; 2002US-0396412P.	
PR	13-AUG-2002; 2002US-0403517P.	
PR	06-SEP-2002; 2002US-00236417.	

(CURA-) CURAGEN CORP.

XX Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyanekar UM, Miller CE, Ooi CE, Ort T, Padigar M, Patrajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twimlow N, Vernet CAM, Voss EZ,
PI Zernusen BD, Zhong M;
XX
XX WPI, 2003-313241/30.
DR P-PSDB; ABR54267.
DR
XX
XX
PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS Claim 20; Page 245; 460pp; English.

XX The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeamic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis
CC valve disease, tuberosus sclerosis, scleroderma, obesity, transplantation
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match	68.4%;	Score 1370;	DB 7;	Length 1650;
Best Local Similarity	92.0%;	Pred. No. 4.1e-259;		
Matches 1509; Conservative	0;	Mismatches 5;	Indels 126;	Gaps 1;

QY	134	GCTCTTTACCATGCGCTGGATCATCTTCCTTGAATGACGAAGCTTGCTGGCCAAAAGATG	193
Db	11	GCTCTTTACCATGCGCTGGATCACTTCCTTGAATGACGAAGCTTGCTGGCCAAAAGATG	70
QY	194	TGGGAATGTGTCGCCCTTGAGATCTATTTTCTCTCAATATGTGTGATCAGCAGAGTTGG	253
Db	71	TTGGGATTTGTGCCCTTGAGATCTATTTTCTCTCAATATGTGTGATCAGCAGAGTTGG	130
QY	254	AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCCTGGGCCAGCCAAAGATGG	313
Db	131	AAAAATATGATGGTGTAGATGCTGGGAAGTATACCATTTGGCCTGGGCCAGCCAAAGATGG	190
QY	314	GCTTCTGCACAGATAGAGAAGATATTAATCTCTTTGCATGACTGTGGTTCAGAACTTAA	373
Db	191	GCTTCTGCACAGATAGAGAAGATATTAATCTCTTTGCATGACTGTGGTTCAGAACTTAA	250
QY	374	TGGAGAGAAATAACCTTTCTCTATGATTTGCATTGGGGCGCTGGAAGTTGGAAACAGAGACAA	433
Db	251	TGGAGAGAAATAACCTTTCTCTATGATTTGCATTGGGGCGCTGGAAGTTGGAAACAGAGACAA	310
QY	434	TCATTCGACCAATCAAAGTCTGTGAAGACTAATTGTATGTCAGCTGTTGAAGAGTCTGGGA	493
Db	311	TCATTCGACCAATCAAAGTCTGTGAAGACTAATTGTATGTCAGCTGTTGAAGAGTCTGGGA	370
QY	494	ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGAGAGCA CAGCTGCTGCT	553
Db	371	ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGAGAGCA CAGCTGCTGCT	430
QY	554	TCATGCTGTTAACTGGAGTTGAGTCAGCTCTTGGGAT-----	591
Db	431	TCATGCTGTTAACTGGAGTTGAGTCAGCTCTTGGGATGGACGGTATGCCCTGTAGTTG	490
QY	592	-----	591
Db	491	CAGGAGATATTGCTGTATATGCCACAGGAATGCTAGACCTACAGGTGGAGTTGAGCAG	550
QY	592	-----GGGCTGCTGGGACAC	607
Db	551	TAGCTCTGCTAATTGGGCCAAATGCTCTTAAATTTTGAACGAGGCGCTTCGTGGGACAC	610
QY	608	ATAATGCAACATGCTATGATTTTAAAGCCGTANATGCTATCTGAATATCCTATAGTAG	667

Dp	611	ATATGCAACATGCTATGATTTTTTACAAAGCCGTGATATGCTATCTGAATATCCTATAGTAG	670
QY	668	ATGAAAACTCTCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT	727
Dp	671	ATGGGAACTCTCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTATCTGTCTACT	730
QY	728	GCAAAAAAGATCCATGCCAGTGGCAGAAAAGGGAATGATTAAGATTTTACCTTGAATG	787
Dp	731	GCAAAAAGATCCATGCCAGTGGCAGAAAAGGGAATGATTAAGATTTTACCTTGAATG	790
QY	788	ATTTGGCTTCATGATCTTTCACTCACCATATTTGTAACTGGTTCAGAAATCTCTAGCTC	847
Dp	791	ATTTGGCTTCATGATCTTTCACTCACCATATTTGTAACTGGTTCAGAAATCTCTAGCTC	850
QY	848	GGATGTTGCTGAATGACTCTCTTAATGACCAGATAAGATAAAAAATGATCTATAGTG	907
Dp	851	GGATGTTGCTGAATGACTCTCTTAATGACCAGATAAGATAAAAAATGATCTATAGTG	910
QY	908	GCCGTGGAAGCCTTTGGGGAATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA	967
Dp	911	GCCGTGGAAGCCTTTGGGGAATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA	970
QY	968	AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCAGTCAGAAAAACAAGCATCTTACTTG	1027
Dp	971	AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCAGTCAGAAAAACAAGCATCTTACTTG	1030
QY	1028	TATCAATCAAAATGGAATATGTACACATCTTCAGTATATGGTCCCTTGCACTGTTC	1087
Dp	1031	TATCAATCAAAATGGAATATGTACACATCTTCAGTATATGGTCCCTTGCACTGTTC	1090
QY	1088	TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAAATTGGAGTCTTCTTATGGTT	1147
Dp	1091	TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAAATTGGAGTCTTCTTATGGTT	1150
QY	1148	CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTCTG	1207
Dp	1151	CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTCTG	1210
QY	1208	CTCTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTG	1267
Dp	1211	CTCTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTG	1270
QY	1268	GTGTGGCACCAAGATGCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG	1327
Dp	1271	GTGTGGCACCAAGATGCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG	1330
QY	1328	TCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAAGTGTACTTAGTTA	1387
Dp	1331	TCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAAGTGTACTTAGTTA	1390
QY	1388	GGGTGATGAAAAAGCACAGAAGACTTACGCTCGGCGTCCCACTCCAATGATGACACTT	1447
Dp	1391	GGGTGATGAAAAAGCACAGAAGACTTACGCTCGGCGTCCCACTCCAATGATGACACTT	1450
QY	1448	TGGATGAAGAGTAGGACTTGTGCATTCAACATAGCAACTGAGCAATTTCCAAAGCCCTG	1507
Dp	1451	TGGATGAAGAGTAGGACTTGTGCATTCAACATAGCAACTGAGCAATTTCCAAAGCCCTG	1510
QY	1508	CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCAATTAGTA	1567
Dp	1511	CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCAATTAGTA	1570
QY	1568	ATGGGGAACATTAAAGATACCTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT	1627
Dp	1571	ATGGGGAACATTAAAGATACCTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT	1630
QY	1628	GGGGGTATGGGAACAGTTGG 1647	
Dp	1631	GGGGGTATGGGAACAGTTGG 1650	

RESULT 6

ACC62328	
ID	ACC62328 standard; cDNA; 1650 BP.
XX	
AC	ACC62328;
XX	
DT	23-JUN-2003 (first entry)
XX	
DE	Human NOV41c encoding cDNA SEQ ID NO:185.
XX	
KW	Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW	anorectic; immunosuppressive; cytostatic; antidiabetic; antifertility;
KW	haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;
KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma
KW	congenital heart defect; aortic stenosis; valve disease; transplantation
KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW	fertility; haemophilia; hypercoagulation; AIDS; bronchial asthma; anorexia;
KW	idiopathic thrombocytopenic purpura; AIDS; infectious disease; cancer;
KW	Crohn's disease; multiple sclerosis; Parkinson's disease; cancer;
KW	cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW	immune disorder; haematopoietic disorder; dyslipidaemia;
KW	metabolic syndrome X; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO2003023001-A2.
XX	
PD	20-MAR-2003.
XX	
PF	09-SEP-2002; 2002WO-US028538.
XX	
PR	07-SEP-2001; 2001US-0318120P.
PR	07-SEP-2001; 2001US-0318184P.
PR	10-SEP-2001; 2001US-0318430P.
PR	17-SEP-2001; 2001US-0322636P.
PR	17-SEP-2001; 2001US-0322781P.
PR	17-SEP-2001; 2001US-0322816P.
PR	17-SEP-2001; 2001US-0322817P.
PR	19-SEP-2001; 2001US-0323519P.
PR	20-SEP-2001; 2001US-0323631P.
PR	20-SEP-2001; 2001US-0323636P.
PR	25-SEP-2001; 2001US-0324969P.
PR	25-SEP-2001; 2001US-0325091P.
PR	26-SEP-2001; 2001US-0324990P.
PR	14-DEC-2001; 2001US-0341144P.
PR	26-FEB-2002; 2002US-0359599P.
PR	05-MAR-2002; 2002US-0361663P.
PR	03-MAY-2002; 2002US-0377908P.
PR	17-MAY-2002; 2002US-0381483P.
PR	29-MAY-2002; 2002US-0383863P.
PR	02-JUL-2002; 2002US-0393332P.
PR	17-JUL-2002; 2002US-0396412P.
PR	13-AUG-2002; 2002US-0403517P.
PR	06-SEP-2002; 2002US-00236417.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI	Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI	Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI	Gargoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI	Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI	Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI	Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI	Spaderna SK, Spytek KA, Taupier RJ, Twimlow N, Vernet CAM, Voss EZ;
PI	Zerhusen BD, Zhong M;
XX	
DR	WPI: 2003-313241/30.
XX	
DR	P-PSDB; ABR54259.

```

PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS Claim 20; Page 241; 460bp; English.
XX
XX The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaemic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberosus sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match      68.4%; Score 1370; DB 7; Length 1650;
Best Local Similarity 92.0%; Pred. No. 4.1e-259;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

QY 134 GCTCTTCCACCATGCGCTGGATCACTTCCCTTGAATGAGAAGCTTGCTGGCCAAAGATG 193
    |||||
DB 11 GCTCTTCCACCATGCGCTGGATCACTTCCCTTGAATGAGAAGCTTGCTGGCCAAAGATG 70

QY 194 TGGGAATTGTTGCCCTTGAGATCTATTTCCCTTCAATATGTTGATCAAGCAGAGTTGG 253
    |||||
DB 71 TTGGGATTGTTGCCCTTGAGATCTATTTCCCTTCAATATGTTGATCAAGCAGAGTTGG 130

QY 254 AAAAATATGATGCTGTAGATGCTGGAAGTATACCATGGCTTGGCCAGGCCAAGATGG 313
    |||||
DB 131 AAAAATATGATGCTGTAGATGCTGGAAGTATACCATGGCTTGGCCAGGCCAAGATGG 190

QY 314 GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTGCATGACTGTGGTTCAGAATCTTA 373
    |||||
DB 191 GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTGCATGACTGTGGTTCAGAATCTTA 250

QY 374 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAA 433
    |||||
DB 251 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAA 310

QY 434 TCATCGACAAATCAAAGCTCTGTGAAGACTTAATTGATGACAGCTGTTGAAGAGTCTGGGA 493
    |||||
DB 311 TCATCGACAAATCAAAGCTCTGTGAAGACTTAATTGATGACAGCTGTTGAAGAGTCTGGGA 370

QY 494 ATACAGATATAGAAGGAATCGACACACAACTAATGCATGCTATGGAGGCACAGCTGCTGCT 553
    |||||
DB 371 ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGCT 430

QY 554 TCAATGCTGTTAACGTGATTGAGTCCAGCTCTTGGGAT----- 591
    |||||
DB 431 TCAATGCTGTTAACGTGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCCTGTAGTTG 490

QY 592 ----- 591
DB 491 CAGAGATATTGCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGAGCAG 550

QY 592 ----- 591
DB 592 ----- 591
    GGGCTTGGTGGGACAC 607
    |||||

```

Db 551 TAGCTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTGCGGACAC 610
QY 608 ATATGCAACATGCCCTATGATTTTACAAGCCTGATATGCTATCTGAATATCTATAGTAG 667
Db 611 ATATGCAACATGCCCTATGATTTTACAAGCCTGATATGCTATCTGAATATCTATAGTAG 670
QY 668 ATGGAAGACTCTCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTACTGTCTACT 727
Db 671 ATGGGAACCTCCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTACTGTCTACT 730
QY 728 GCAAAAAGATCCATGCCAGTGGCAGAAAGGGAATGATAAAGATTTTACCTTGAATG 787
Db 731 GCAAAAAGATCCATGCCAGTGGCAGAAAGGGAATGATAAAGATTTTACCTTGAATG 790
QY 788 ATTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTTAGCTC 847
Db 791 ATTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTTAGCTC 850
QY 848 GGATGTGCTGAATGACTTCTTAATGACCAAGATAGAGATPAAAAATAGTACTATAGTG 907
Db 851 GGATGTGCTGAATGACTTCTTAATGACCAAGATAGAGATPAAAAATAGTACTATAGTG 910
QY 908 GCCTGGAAGCCTTGGGGGATGTTAATTAGAGACACCTACTTTGATAGAGATGTGAGA 967
Db 911 GCCTGGAAGCCTTGGGGGATGTTAATTAGAGACACCTACTTTGATAGAGATGTGAGA 970
QY 968 AGGCATTATGAAGGCTAGCTCTGAACCTCTTCAGTCAGAAAACAAAGGCATCTTACTTG 1027
Db 971 AGGCATTATGAAGGCTAGCTCTGAACCTCTTCAGTCAGAAAACAAAGGCATCTTACTTG 1030
QY 1028 TATCAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTGCACTGTTC 1087
Db 1031 TATCAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTGCACTGTTC 1090
QY 1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAAATTGGAGTGTCTTATAGTT 1147
Db 1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAAATTGGAGTGTCTTATAGTT 1150
QY 1148 CTGTTTGGCTGCACACTCTGTACTCTCTTAAAGTCACACAAGATGTACACCGGGGCTCG 1207
Db 1151 CTGTTTGGCTGCACACTCTGTACTCTCTTAAAGTCACACAAGATGTACACCGGGGCTCG 1210
QY 1208 CTCTTGATPAAATPACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1267
Db 1211 CTCTTGATPAAATPACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1270
QY 1268 GTGTGGCACAGATGTCTTCGCTGAAAACATGAAAGCTCAGAGAGGACACCCATCATTTGG 1327
Db 1271 GTGTGGCACAGATGTCTTCGCTGAAAACATGAAAGCTCAGAGAGGACACCCATCATTTGG 1330
QY 1328 TCAACTATATTTCCCGAGGTTCAATAGATTCACTCTTGAAGGAACGTGGTACTTAGTGA 1387
Db 1331 TCAACTATATTTCCCGAGGTTCAATAGATTCACTCTTGAAGGAACGTGGTACTTAGTGA 1390
QY 1388 GGGTGATGAAAAGCACAGAGAAGTACGCTCGGCGTCCCACTCCAATGATGACACTT 1447
Db 1391 GGGTGATGAAAAGCACAGAGAAGTACGCTCGGCGTCCCACTCCAATGATGACACTT 1450
QY 1448 TGGATGAAGAGTAGGACTTGTGCAATTCAACATAGCAACTGAGCATATTCCAAGCCCTG 1507
Db 1451 TGGATGAAGAGTAGGACTTGTGCAATTCAACATAGCAACTGAGCATATTCCAAGCCCTG 1510
QY 1508 CCAAGAAAAGTACCAAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAGCTGTCTATTAGTA 1567
Db 1511 CCAAGAAAAGTACCAAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAGCTGTCTATTAGTA 1570
QY 1568 ATGGGAACAVTAAGATACTGTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT 1627
Db 1571 ATGGGAACAVTAAGATACTGTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT 1630
QY 1628 GGGGGTATGGGAACAGTTGG 1647
Db 1631 GGGGGTATGGGAACAGTTGG 1650

RESULT 7
ACCE2326
ID ACC62326 standard, cDNA; 1650 BP.
XX
AC ACC62326;
XX
DT 23-JUN-2003 (first entry)
XX
DE Human NOV41a encoding cDNA SEQ ID NO:181.
XX
KW Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cyostatic; antidiabetic; antiinfertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003023001-A2.
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028538.
XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0393363P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli E, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Pattnarajan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zerhusen BD, Zhong M;
XX
DR MPI; 2003-313241/30.
DR P-PSDB; ABR54257.

XX Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS Claim 20; Page 240; 460pp; English.
XX
CC The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiatic,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeamic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tubercous sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match 68.4%; Score 1370; DB 7; Length 1650;
Best Local Similarity 92.0%; Pred. No. 4.1e-259;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

QY 134 GCTCTTTCACCATGCGCTGGATCACTTCTTGAATGCAGAAGCTTGCCCAAAAGATG 193
Db 11 GCTCTTTCACCATGCGCTGGATCACTTCTTGAATGCAGAAGCTTGCCCAAAAGATG 70
QY 194 TGGGAATGTTGCCCTTGAGATCTATTTCTCTCATATGTTGATCAAGCAGAGTTGG 253
Db 71 TTGGGATTTGTCCTTGAGATCTATTTCTCTCATATGTTGATCAAGCAGAGTTGG 130
QY 254 AAAAATATGATGCTGTAGATGCTGGAAGATATACCATTTGGCTTGGCCAGCCAAAGATGG 313
Db 131 AAAAATATGATGCTGTAGATGCTGGAAGATATACCATTTGGCTTGGCCAGCCAAAGATGG 190
QY 314 GCTTTCGACAGATAGAGAATATTACTCTTTGCATGACTGTGTTCAGAACTTTA 373
Db 191 GCTTTCGACAGATAGAGAATATTACTCTTTGCATGACTGTGTTCAGAACTTTA 250
QY 374 TGGAGAAGAAATAACCTTTCTATGATTCATTTGGCGGCTGGAAGTTGGAACAGAGACAA 433
Db 251 TGGAGAAGAAATAACCTTTCTATGATTCATTTGGCGGCTGGAAGTTGGAACAGAGACAA 310
QY 434 TCATCGACAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGTCTGGGA 493
Db 311 TCATCGACAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGTCTGGGA 370
QY 494 ATACAGATATAGAAGGAATCGACACAATAATGCTATGAGGACACAGCTGCTGCT 553
Db 371 ATACAGATATAGAAGGAATCGACACAATAATGCTATGAGGACACAGCTGCTGCT 430
QY 554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGAT----- 591
Db 431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGAGCGGTATGCCCTGTAGTTG 490
QY 592 ----- 591
Db 491 CAGAGATATTTGCTGTATATGCACAGGAATGCTAGACCTACAGGTGAGTTGAGACAG 550

QY 592 -----GGGCTTCGTGGACAC 607
Db 551 TAGCTTCCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGACAC 610
QY 608 ATATGACACATGCGCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 667
Db 611 ATATGCAACATGCGCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 670
QY 668 ATGGAAACTCTCCATPACAGTGTACTCCTCAGTGCATTGACCGGCTGCTACTCTGTCTACT 727
Db 671 ATGGGAACTCTCCATPACAGTGTACTCCTCAGTGCATTGACCGGCTGCTACTCTGTCTACT 730
QY 728 GCAAAAGATCCATGCCCCAGTGGCAGAAAGAGGAAATGATAAAGATTTTACCTTGAATG 787
Db 731 GCAAAAGATCCATGCCCCAGTGGCAGAAAGAGGAAATGATAAAGATTTTACCTTGAATG 790
QY 788 ATTTGCGCTTCATGATCTTTCACTCAACCATATTTGTAACCTGTTCAAGAAATCTAGCTC 847
Db 791 ATTTGCGCTTCATGATCTTTCACTCAACCATATTTGTAACCTGTTCAAGAAATCTAGCTC 850
QY 848 GGATGTTGCTGAATGACTTCTTAAATGACCAGAAATAGAGATAAATAGTATCTATAGTG 907
Db 851 GGATGTTGCTGAATGACTTCTTAAATGACCAGAAATAGAGATAAATAGTATCTATAGTG 910
QY 908 GCGTGAAACCTTTGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA 967
Db 911 GCGTGAAACCTTTGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA 970
QY 968 AGGCATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAGGCATCTTTACTTG 1027
Db 971 AGGCATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAGGCATCTTTACTTG 1030
QY 1028 TATCAATCAAAATGGAATATGTAACACATCTTCAGTATATGTTCCCTTGCACTGTTC 1087
Db 1031 TATCAATCAAAATGGAATATGTAACACATCTTCAGTATATGTTCCCTTGCACTGTTC 1090
QY 1088 TAGCAGATACTCACCTCAGCAATTAAGCAGGAAGAAATGGAGTGTTCCTATGCTT 1147
Db 1091 TAGCAGATACTCACCTCAGCAATTAAGCAGGAAGAAATGGAGTGTTCCTATGCTT 1150
QY 1148 CTGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACGGGGTCTG 1207
Db 1151 CTGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACGGGGTCTG 1210
QY 1208 CTCTGATTAATAATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAAGAACTG 1267
Db 1211 CTCTGATTAATAATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAAGAACTG 1270
QY 1268 GTGTGGACCCAGATGTCTTCGCTGAAGAAACATGAAGCTCAGAGAGACACCCATCATTTGG 1327
Db 1271 GTGTGGACCCAGATGTCTTCGCTGAAGAAACATGAAGCTCAGAGAGACACCCATCATTTGG 1330
QY 1328 TCACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTACTTAGTTA 1387
Db 1331 TCACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTACTTAGTTA 1390
QY 1388 GGGTGAATGAAAAAGCACAGAAAGAACTTAAGCTCGGCGTCCCACTCCAATGATGACACTT 1447
Db 1391 GGGTGAATGAAAAAGCACAGAAAGAACTTAAGCTCGGCGTCCCACTCCAATGATGACACTT 1450
QY 1448 TGGATGAAGAGTAGAGACTTGTGCATTCAAAACATAGCAACTGAGCATATTTCCAAGCCCTG 1507
Db 1451 TGGATGAAGAGTAGAGACTTGTGCATTCAAAACATAGCAACTGAGCATATTTCCAAGCCCTG 1510
QY 1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACTGAAGCAGCTGTCAATTAGTA 1567
Db 1511 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACTGAAGCAGCTGTCAATTAGTA 1570
QY 1568 ATGGGAAACATTAAGATACTCTGTGAGGTGCAAGACTTCAAGGCTGGGTGGCATGGGGT 1627
Db 1571 ATGGGAAACATTAAGATACTCTGTGAGGTGCAAGACTTCAAGGCTGGGTGGCATGGGGT 1630

QY 1628 GGGGGTATGGAAcAGTTGG 1647
|||||
Db 1631 GGGGGTATGGAAcAGTTGG 1650
RESULT 8
ACc62330
ID ACC62330 standard; cDNA; 1650 BP.
XX
AC ACC62330;
XX
DT 23-JUN-2003 (first entry)
XX
DE Human NOV41e encoding cDNA SEQ ID NO:189.
XX
KW Human; NOvX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cyostatic; antidiabetic; antiinfertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeamic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tubercous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003023001-A2.
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028538.
XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 26-SEP-2001; 2001US-0341144P.
PR 26-SEP-2001; 2001US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
XX
PA (CURA-) CURAGEN CORP.
XX
PI Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catercon E, Chant JS, Chaudhuri A;
PI Crabtree J, DiPippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyanekar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patlurajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA;
PI Spaderna SK, Splytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zerhusen BD, Zhong M;

XX
DR WPI; 2003-313241/30.
DR P-PSDB; ABR54261.
XX
PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS Claim 20; Page 242; 460pp; English.
XX
CC The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62336 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cyostatic,
CC antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeamic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tubercous sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;
XX
Query Match 68.4%; Score 1370; DB 7; Length 1650;
Best Local Similarity 92.0%; Pred. No. 4.1e-259;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;
QY 134 GCTCTTTCACCATGCGCTGATCACTTCCCTTGAATGACAGAGCTGCGCCAAAAGATG 193
|||
Db 11 GCTCTTTCACCATGCGCTGATCACTTCCCTTGAATGACAGAGCTGCGCCAAAAGATG 70
QY 194 TGGGAATGTGGCCCTTGAGATCTATTTCCCTCTCAATATGTTGATCAAGCAGAGTTGG 253
|||
Db 71 TTGGGATGTGGCCCTTGAGATCTATTTCCCTCTCAATATGTTGATCAAGCAGAGTTGG 130
QY 254 AAAAATATGATGTTAGATGCTGGAAGATATACCATTTGGCTTGGGCCAGGCCAAGATGG 313
|||
Db 131 AAAAATATGATGTTAGATGCTGGAAGATATACCATTTGGCTTGGGCCAGGCCAAGATGG 190
QY 314 GCTTCTGACAGATAGAGAAGATATTAATCTCTTTGCATGACTGTGTTCAGATCTTA 373
|||
Db 191 GCTTCTGACAGATAGAGAAGATATTAATCTCTTTGCATGACTGTGTTCAGATCTTA 250
QY 374 TGGAGAGAAATTAACCTTTCTATGATTGCAATTGGCGGCTGGAAATTGGAACAGACAA 433
|||
Db 251 TGGAGAGAAATTAACCTTTCTATGATTGCAATTGGCGGCTGGAAATTGGAACAGACAA 310
QY 434 TCATCGACAAATCAAAAGTCTGTGAAGACTAATTTGATGCAAGCTGTGTTGAAGAGTCTGGGA 493
|||
Db 311 TCATCGACAAATCAAAAGTCTGTGAAGACTAATTTGATGCAAGCTGTGTTGAAGAGTCTGGGA 370
QY 494 ATACAGATATAGAAAGGAATCGACACAACATAATGCAATGCTATGAGAGGCACAGCTGCTTCT 553
|||
Db 371 ATACAGATATAGAAAGGAATCGACACAACATAATGCAATGCTATGAGAGGCACAGCTGCTTCT 430
QY 554 TCAATGCTGTAACTGGAATGCCAGCTCTTGGAT----- 591
|||
Db 431 TCAATGCTGTAACTGGAATGCCAGCTCTTGGATGACGGATATGCCCTGTAGTTG 490

QY 592 ----- 591
Db 491 CAGAGATATTGCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTGGAGCAG 550
QY 592 -----GGGCTTGGTGACAC 607
Db 551 TAGCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGACAC 610
QY 608 ATATGCACATGCGCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 667
Db 611 ATATGCACATGCGCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 670
QY 668 ATGGAAAACTCTCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT 727
Db 671 ATGGAAAACTCTCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT 730
QY 728 GCAAAAAAGATCCATGCCCCAGTGGCAGAAAAAGGGAAATGATAAGATTTTACCCTTGAATG 787
Db 731 GCAAAAAAGATCCATGCCCCAGTGGCAGAAAAAGGGAAATGATAAGATTTTACCCTTGAATG 790
QY 788 ATTTGGCTTCATGATCTTTCACTCAACCATATTGTAACTGGTTCAGAAATCTCTAGCTC 847
Db 791 ATTTGGCTTCATGATCTTTCACTCAACCATATTGTAACTGGTTCAGAAATCTCTAGCTC 850
QY 848 GGAATGTTGCTGAATGACTTCCTTAATGACCAGAAATAGAGATAAAATAGTATCTATAGTG 907
Db 851 GGAATGTTGCTGAATGACTTCCTTAATGACCAGAAATAGAGATAAAATAGTATCTATAGTG 910
QY 908 GCGTGGAAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA 967
Db 911 GCGTGGAAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA 970
QY 968 AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCAGTCAGAAAACAAAGGCATCTTTACTTG 1027
Db 971 AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCAGTCAGAAAACAAAGGCATCTTTACTTG 1030
QY 1028 TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTC 1087
Db 1031 TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTC 1090
QY 1088 TAGCAGACTACTCACCTCAGCAATTAGCAGGGAAGAAATTGGAGTGTCTTATAGTGT 1147
Db 1091 TAGCAGACTACTCACCTCAGCAATTAGCAGGGAAGAAATTGGAGTGTCTTATAGTGT 1150
QY 1148 CTGGTTTGGCTGCCACTCTGTACTCTTTAAAGTCACACAAAGATGCTACACCGGGGTCTG 1207
Db 1151 CTGGTTTGGCTGCCACTCTGTACTCTTTAAAGTCACACAAAGATGCTACACCGGGGTCTG 1210
QY 1208 CTCTTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAAGCTTGATTCAGAAGCTG 1267
Db 1211 CTCTTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAAGCTTGATTCAGAAGCTG 1270
QY 1268 GTGTGGCACCAAGATGTCTTGGCTGAAAAACATGAAGCTCAGAGAGACACCCATCATTTGG 1327
Db 1271 GTGTGGCACCAAGATGTCTTGGCTGAAAAACATGAAGCTCAGAGAGACACCCATCATTTGG 1330
QY 1328 TCAACTATATTTCCCGAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTACTTAGTTA 1387
Db 1331 TCAACTATATTTCCCGAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTACTTAGTTA 1390
QY 1388 GGGTGAATGAAAAAGCACAGAAAGACTTACGCTCGGGCTCCCACTCCAATGATGACACTT 1447
Db 1391 GGGTGAATGAAAAAGCACAGAAAGACTTACGCTCGGGCTCCCACTCCAATGATGACACTT 1450
QY 1448 TGGATGAAGGAGTAGGACTGTGCAATTCAACATAGCAACTGAGCATATTTCCAAGCCCTG 1507
Db 1451 TGGATGAAGGAGTAGGACTGTGCAATTCAACATAGCAACTGAGCATATTTCCAAGCCCTG 1510
QY 1508 CCAAGAAAGTACCAAGACTCCCTGCACAGCAGCAGAACTGGAAGCAGCTGTCAATAGTA 1567
Db 1511 CCAAGAAAGTACCAAGACTCCCTGCACAGCAGCAGAACTGGAAGCAGCTGTCAATAGTA 1570
QY 1568 ATGGGAACATTAAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGCATGGGGT 1627

Db 1571 ATGGGAACATTAAAGTACTCTGTGAGGTGCAAGACTTCAGGGTGGGGCATGGGGT 1630
QY 1628 GGGGGTATGGGAACAGTTGG 1647
Db 1631 GGGGGTATGGGAACAGTTGG 1650
RESULT 9
ACC62338
ID ACC62338 standard; cDNA; 1650 BP.
XX
AC ACC62338;
DT 23-JUN-2003 (first entry)
XX
DE Human NOV41m encoding cDNA SEQ ID NO:205.
XX
KW Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003023001-A2.
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028538.
XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
PA
XX
PA (CURA-) CURAGEN CORP.
PI Agee ML, Alsebrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Paturajan M;

PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Voos EZ,
PI Spaderna SK, Spyrek KA, Taupier RJ, Twomlow N, Vernet CAM,
PI Zernusen BD, Zhong M,
XX WPI; 2003-313241/30.
DR P-PSDB; ABR54269.

XX Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.

PS Claim 20; Page 246; 460pp; English.

XX The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antiinfectility, haemostatic, antiinflammatory, anti-HIV,
CC antiassthetic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeimic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valvular disease, tuberosus sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match 68.4%; Score 1370; DB 7; Length 1650;
Best Local Similarity 92.0%; Pred. No. 4.1e-259;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

QY 134 GCTCTTACCATGCTGATCACTTCTTGAATGCAGAACTTGCTGCCAAAAGATG 193
DB 11 GCTCTTACCATGCTGATCACTTCTTGAATGCAGAACTTGCTGCCAAAAGATG 70
QY 194 TGGGAATGTTGCCCTTGAGATCTATTTCTTCAATATGTTGATCAAGCAGAGTTGG 253
DB 71 TTGGGATGTTGCCCTTGAGATCTATTTCTTCAATATGTTGATCAAGCAGAGTTGG 130
QY 254 AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCTTGGGCCAAGCAGATGG 313
DB 131 AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCTTGGGCCAAGCAGATGG 190
QY 314 GCTTCTGCACAGATAGAGAATATTAACTCTCTTGCATGACTGTGTTCAAACTTTA 373
DB 191 GCTTCTGCACAGATAGAGAATATTAACTCTCTTGCATGACTGTGTTCAAACTTTA 250
QY 374 TGGAGAGAATAAAGCTTCCATGATTTGCATTTGGCGCGCTGGAAGTTGGAACAGACAA 433
DB 251 TGGAGAGAATAAAGCTTCCATGATTTGCATTTGGCGCGCTGGAAGTTGGAACAGACAA 310
QY 434 TCATCGACAATCAAAGCTGTGAAGACTAATTTGATGCAGCTGTTGAAGAGCTGGGA 493
DB 311 TCATCGACAATCAAAGCTGTGAAGACTAATTTGATGCAGCTGTTGAAGAGCTGGGA 370
QY 494 ATACAGATATAGAAAGAAATCGACAACTAATGATGCTATGAGGACACAGCTGCTGTCT 553
DB 371 ATACAGATATAGAAAGAAATCGACAACTAATGATGCTATGAGGACACAGCTGCTGTCT 430
QY 554 TCAATGCTGTTAACTGATTTGAGTCCAGCTCTTGGGAT----- 591

DB 431 TCAATGCTGTTAACTGATTTGAGTCCAGCTCTTGGGATGACGGTATGCCCTGTAGATTG 490
QY 592 ----- 591
DB 491 CAGAGATATTGCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGACAG 550
QY 592 -----GGGCTTCGTGGACAC 607
DB 551 TAGCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGACAC 610
QY 608 ATATGCAACATGCTATGATTTTTCAAAGCCTGATATGCTATCTGAATATCTTAAGTAG 667
DB 611 ATATGCAACATGCTATGATTTTTCAAAGCCTGATATGCTATCTGAATATCTTAAGTAG 670
QY 668 ATGAAAACCTCTCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTACTCTGTACT 727
DB 671 ATGGAAACTCTCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTACTCTGTACT 730
QY 728 GCAAAAAGATCCATGCCAGTGGCAGAAAAGGGAATGATTAAGATTTTACCCTTGAATG 787
DB 731 GCAAAAAGATCCATGCCAGTGGCAGAAAAGGGAATGATTAAGATTTTACCCTTGAATG 790
QY 788 ATTTGGCTTCATGATCTTTCACTCACCATTGTTAACTGGTTCAGAAATCTCTAGCTC 847
DB 791 ATTTGGCTTCATGATCTTTCACTCACCATTGTTAACTGGTTCAGAAATCTCTAGCTC 850
QY 848 GGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAGATTAATAATAGTATCTAAGTG 907
DB 851 GGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAGATTAATAATAGTATCTAAGTG 910
QY 908 GCCTGGAAGCCTTTGGGATGTTAAATTAGAGACACCTACTTTGATAGAGATGGAGA 967
DB 911 GCCTGGAAGCCTTTGGGATGTTAAATTAGAGACACCTACTTTGATAGAGATGGAGA 970
QY 968 AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCACTCAGTCAAGAAAACAAAGGCATCTTACTTG 1027
DB 971 AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCACTCAGTCAAGAAAACAAAGGCATCTTACTTG 1030
QY 1028 TATCAAAATCAAAATGGAATATGTACACATCTTCACTATATGTTCCCTTGCACTGTTC 1087
DB 1031 TATCAAAATCAAAATGGAATATGTACACATCTTCACTATATGTTCCCTTGCACTGTTC 1090
QY 1088 TAGCACAGTACTCACTCAGCAATTAAGCAGGGAAGAAATTGGAGTGTTCCTATAGGTT 1147
DB 1091 TAGCACAGTACTCACTCAGCAATTAAGCAGGGAAGAAATTGGAGTGTTCCTATAGGTT 1150
QY 1148 CTGCTTGGCTGCCACTCTGTACTCTCTTAAGTCAACAAGATGCTACACCGGGGCTCG 1207
DB 1151 CTGCTTGGCTGCCACTCTGTACTCTCTTAAGTCAACAAGATGCTACACCGGGGCTCG 1210
QY 1208 CTCTTGATAAATAACAGCAAGTTATGTGATCTTAAATCAAGCCTTGATTCAGAACTG 1267
DB 1211 CTCTTGATAAATAACAGCAAGTTATGTGATCTTAAATCAAGCCTTGATTCAGAACTG 1270
QY 1268 GTGTGGCACCAGATGTCTTGGCTGAAAACATGAAGCTCAGAGAGACACCCATCAATTGG 1327
DB 1271 GTGTGGCACCAGATGTCTTGGCTGAAAACATGAAGCTCAGAGAGACACCCATCAATTGG 1330
QY 1328 TCAACTATATTTCCCGAGGTTCAATGATTTCACTCTTTGAAGGAACGTGTACTTAGTTA 1387
DB 1331 TCAACTATATTTCCCGAGGTTCAATGATTTCACTCTTTGAAGGAACGTGTACTTAGTTA 1390
QY 1388 GGGTGGATGAAAAGCACAGAGAATCTTACGCTGGCGCTCCCATCTCCAATGATGACACTT 1447
DB 1391 GGGTGGATGAAAAGCACAGAGAATCTTACGCTGGCGCTCCCATCTCCAATGATGACACTT 1450
QY 1448 TGGATGAAGGAGTAGGACTTTGTGCATTCAAAACATAGCAACTGAGCATATTTCCAAGCCCTG 1507
DB 1451 TGGATGAAGGAGTAGGACTTTGTGCATTCAAAACATAGCAACTGAGCATATTTCCAAGCCCTG 1510
QY 1508 CCAAGAAAGTACCAAGACTCCCTGCGCACAGCAGAGAACCTGAAGCAGACTGTCAATTAGTA 1567

Db	1511	CCAAGAAAGTACCAAGACTCCCTGCCACACAGACAGAACTGAAGCAGCTGTCAATTA	1570
Qy	1568	ATGGGGAACATTAGATCTCTGTGAGGTGCAAGACTTCAGGGTGGGCATGGGGT	1627
Db	1571	ATGGGGAACATTAGATCTCTGTGAGGTGCAAGACTTCAGGGTGGGCATGGGGT	1630
Qy	1628	GGGGGTATGGGAACAGTTGG	1647
Db	1631	GGGGGTATGGGAACAGTTGG	1650
RESULT 10			
ACC62334			
ID	ACC62334	standard; cdNA; 1650 BP.	
XX	AC	ACC62334;	
XX	DT	23-JUN-2003 (first entry)	
XX	DE	Human NOV411 encoding cdNA SEQ ID NO:197.	
XX	KW	Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;	
KW	anorectic; immunosuppressive; cyostatic; antidiabetic; antifertility;		
KW	haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;		
KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;		
KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;		
KW	congenital heart defect; aortic stenosis; valve disease; transplantation;		
KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;		
KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;		
KW	fertility; haemophilia; hypercoagulation; graft versus host disease;		
KW	idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;		
KW	Crohn's disease; multiple sclerosis; infectious disease; cancer;		
KW	cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;		
KW	immune disorder; haematopoietic disorder; dyslipidaemia;		
KW	metabolic syndrome X; gene; ss.		
XX	OS	Homo sapiens.	
XX	PN	WO2003023001-A2.	
XX	PD	20-MAR-2003.	
XX	PF	09-SEP-2002; 2002WO-US028538.	
XX	PR	07-SEP-2001; 2001US-0318120P.	
PR	07-SEP-2001; 2001US-0318184P.		
PR	10-SEP-2001; 2001US-0318430P.		
PR	17-SEP-2001; 2001US-0322636P.		
PR	17-SEP-2001; 2001US-0322781P.		
PR	17-SEP-2001; 2001US-0322816P.		
PR	17-SEP-2001; 2001US-0322817P.		
PR	19-SEP-2001; 2001US-0323519P.		
PR	20-SEP-2001; 2001US-0323631P.		
PR	20-SEP-2001; 2001US-0323636P.		
PR	25-SEP-2001; 2001US-0324969P.		
PR	25-SEP-2001; 2001US-0325091P.		
PR	26-SEP-2001; 2001US-0324990P.		
PR	14-DEC-2001; 2001US-0341144P.		
PR	26-FEB-2002; 2002US-0359599P.		
PR	05-MAR-2002; 2002US-0361663P.		
PR	03-MAY-2002; 2002US-0377908P.		
PR	17-MAY-2002; 2002US-0381483P.		
PR	29-MAY-2002; 2002US-0383863P.		
PR	02-JUL-2002; 2002US-0393332P.		
PR	17-JUL-2002; 2002US-0396412P.		
PR	13-AUG-2002; 2002US-0403517P.		
PR	06-SEP-2002; 2002US-00236417.		
XX	PA	(CURA-) CURAGEN CORP.	
XX	PI	Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;	
PI	Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;		
PI	Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;		

PI	Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;	
PI	Kekuda R, Khrantsov NV, Leach MD, Lopley DM, Li L, Liu X;	
PI	Malayanakar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Paturajan M;	
PI	Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;	
PI	Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;	
PI	Zerhusen BD, Zhong M;	
XX	DR	WPI; 2003-313241/30.
DR	P-PSDB; ABR54265.	
XX	PT	Novel human proteins and nucleic acid encoding the proteins, useful for
PT	diagnosis, treatment and prevention of disorders involving the human	
PT	protein or nucleic acid e.g. cardiac and neurological disorders.	
XX	PS	Claim 20; Page 244; 460pp; English.
XX	CC	The present invention describes isolated human NOVX proteins, where X is
CC	1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in	
CC	CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,	
CC	hypotensive, dermatological, anorectic, immunosuppressive, cyostatic,	
CC	antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,	
CC	antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,	
CC	antiparkinsonian and antilipemic activities, and can be used in gene	
CC	therapy. NOVX proteins are useful for treating or preventing a pathology	
CC	associated with a NOVX protein in humans and for treating a syndrome	
CC	associated with the human disease. NOVX nucleic acids, proteins and	
CC	antibodies can be used in the treatment and diagnosis of cardiomyopathy,	
CC	atherosclerosis, hypertension, congenital heart defects, aortic stenosis,	
CC	valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,	
CC	congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic	
CC	disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,	
CC	hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host	
CC	disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,	
CC	infectious disease, anorexia, cancer-associated cachexia, cancer,	
CC	Alzheimer's disease, Parkinson's disease, immune disorders,	
CC	haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.	
CC	ACC62346 to ACC62465 represent PCR primers and probes for human NOVX	
CC	sequences, which are used in examples from the present invention.	
CC	ABR54277 represents a human trypsinogen protein given in comparison with	
CC	the human NOV35b protein in the exemplification of the present invention	
XX	SQ	Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;
Query Match		
Best Local Similarity 68.4%; Score 1370; DB 7; Length 1650;		
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;		
Qy	134	GCTCTTCCACCATGCGCTGGATCACTTCTTGAATGCAGAACTTGGCCAAAGATG 193
Db	11	GCTCTTCCACCATGCGCTGGATCACTTCTTGAATGCAGAACTTGGCCAAAGATG 70
Qy	194	TGGGAATTGTTGCCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGAGTTGG 253
Db	71	TTGGGATTTGTTGCCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGAGTTGG 130
Qy	254	AAAAATATGATGTTGTAGATGCTGGAAGTATACCATGGGCTTGGCCAGGCCAAGATGG 313
Db	131	AAAAATATGATGTTGTAGATGCTGGAAGTATACCATGGGCTTGGCCAGGCCAAGATGG 190
Qy	314	GCTTCTGCACAGATAGAGAAATATTAATCTCTTGGATGACTGTGTTCAAGATCTTA 373
Db	191	GCTTCTGCACAGATAGAGAAATATTAATCTCTTGGATGACTGTGTTCAAGATCTTA 250
Qy	374	TGGAGAGAATAAATCACTTCTATGATTTGCATTGGCGGCTGGAAGTTGGAACAGAGCAA 433
Db	251	TGGAGAGAATAAATCACTTCTATGATTTGCATTGGCGGCTGGAAGTTGGAACAGAGCAA 310
Qy	434	TCATCGACAATCAAAAGTCTGGAAGCTAATTTGATGCAGCTGTTGAAGAGTCTGGGA 493
Db	311	TCATCGACAATCAAAAGTCTGGAAGCTAATTTGATGCAGCTGTTGAAGAGTCTGGGA 370
Qy	494	ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGCT 553

Db 371 ATACAGATATAGAAAGNAATCGACACAACATAATGCATGCTATGAGGCGACAGCTGCTGTCT 430
QY 554 TCAATGCTGTTAATCTGATTGAGTCCAGCTCTTGGGAT----- 591
Db 431 TCAATGCTGTTAATCTGATTGAGTCCAGCTCTTGGGATGAGCGGTATGCCCTGGTAGTTG 490
QY 592 ----- 591
Db 491 CAGGAGATATTGCTGTATATGCCACAGGAATGCTAGACCTACAGGTGGAGTGGAGCAG 550
QY 592 -----GGGCTTCGTGGGACAC 607
Db 551 TAGCTCTGCTAATTGGGCCAAATGCTCCTTAAATTTTGAACGAGGGCTTCGTGGGACAC 610
QY 608 ATATGCACATGCGCTATGATTTTTCACAGCCGTGATATGCTATCTGAATATCCTATAGTAG 667
Db 611 ATATGCACATGCGCTATGATTTTTCACAGCCGTGATATGCTATCTGAATATCCTATAGTAG 670
QY 668 ATGGAAAACTCTCCATACAGTGTCTACCTCAGTGCATTAGACCGGCTGCTACTCTGTCTACT 727
Db 671 ATGGGAAACTCTCCATACAGTGTCTACCTCAGTGCATTAGACCGGCTGCTACTCTGTCTACT 730
QY 728 GCAAAAAAGATCCATGCCCCAGTGGCAGAAAAGAGGAAATGATAAAGATTTTACCTTGATG 787
Db 731 GCAAAAAAGATCCATGCCCCAGTGGCAGAAAAGAGGAAATGATAAAGATTTTACCTTGATG 790
QY 788 ATTTGGCTTCATGATCTTTTCACTCCACCATATTGTTAACTGGTTCAGAAATCTCTAGCTC 847
Db 791 ATTTGGCTTCATGATCTTTCACTCCACCATATTGTTAACTGGTTCAGAAATCTCTAGCTC 850
QY 848 GGATGTTGCTGAATGACTTCTCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTG 907
Db 851 GGATGTTGCTGAATGACTTCTCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTG 910
QY 908 GCCTGGAAGCCTTTGGGGATGTTAAATTAAGAAGACACCTACTTTGATAGAGATGTGGAGA 967
Db 911 GCCTGGAAGCCTTTGGGGATGTTAAATTAAGAAGACACCTACTTTGATAGAGATGTGGAGA 970
QY 968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAAGTCAAGAAACAAAGGCATCTTTACTTG 1027
Db 971 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAAGTCAAGAAACAAAGGCATCTTTACTTG 1030
QY 1028 TATCAAAATCAAAATGAAATATGTACACATCTTCAGTATATAGTTCCTTCATCTGTTC 1087
Db 1031 TATCAAAATCAAAATGAAATATGTACACATCTTCAGTATATAGTTCCTTCATCTGTTC 1090
QY 1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAAATGAGTGTTCCTTATAGTTC 1147
Db 1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAAATGAGTGTTCCTTATAGTTC 1150
QY 1148 CTGTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAGATGCTACACCGGGGCTCTG 1207
Db 1151 CTGTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAGATGCTACACCGGGGCTCTG 1210
QY 1208 CTCTTGATTAATAATACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGTCTG 1267
Db 1211 CTCTTGATTAATAATACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGTCTG 1270
QY 1268 GTGTGGCACCAAGATGTCTTCGCTGAAGAAATGAAGCTCAGAGAGGACACCCCATCATTTGG 1327
Db 1271 GTGTGGCACCAAGATGTCTTCGCTGAAGAAATGAAGCTCAGAGAGGACACCCCATCATTTGG 1330
QY 1328 TCAACTATATTCGCCAGGGTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA 1387
Db 1331 TCAACTATATTCGCCAGGGTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA 1390
QY 1388 GGGTGGATGAAAAGCACAGAGAAGAACTTACGCTGGCGTCCCACTCCAAATGATGACACTT 1447
Db 1391 GGGTGGATGAAAAGCACAGAGAAGAACTTACGCTGGCGTCCCACTCCAAATGATGACACTT 1450
QY 1448 TGGATGAAGAGATAGGACTGTGCAATTCAAACATAGCAACTGAGCATATTTCCAAAGCCCTG 1507
Db 1451 TGGATGAAGAGATAGGACTGTGCAATTCAAACATAGCAACTGAGCATATTTCCAAAGCCCTG 1510

QY 1508 CCAAGAAAAGTACCAAGACTCCCTGCCACACAGCAGACGAACCTGAAAGCAGCTGTCAATTA 1567
Db 1511 CCAAGAAAAGTACCAAGACTCCCTGCCACACAGCAGACGAACCTGAAAGCAGCTGTCAATTA 1570
QY 1568 ATGGGGACATTAAGATACTCTGTGTGAGGTGCAGAGACTTCAGGGTGGGGTGGCATGGGGT 1627
Db 1571 ATGGGGACATTAAGATACTCTGTGTGAGGTGCAGAGACTTCAGGGTGGGGTGGCATGGGGT 1630
QY 1628 GGGGCTATGGGAACAGTTGG 1647
Db 1631 GGGGCTATGGGAACAGTTGG 1650
RESULT 11
ACC62327
ID ACC62327 standard; cDNA, 1650 BP.
XX
AC ACC62327;
DT 23-JUN-2003 (first entry)
XX
DE Human NOV41b encoding cDNA SEQ ID NO:183.
XX
KW Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cytostatic; antidiabetic; antinfertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tubercous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; hematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X; gene; ss.
XX
OS Homo sapiens.
XX
XX WO2003023001-A2.
PN
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028538.
XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
PA (CURA-) CURAGEN CORP.
XX

PI Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patraujan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zerhusen BD, Zhong M;

DR WPI; 2003-313241/30.

DR P-PSDB; ABR54258.

PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.

PS Claim 20; Page 240-241; 460pp; English.

The present invention describes isolated human NOVX proteins, where X is 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiac, hypotensive, dermatological, anorectic, immunosuppressive, cytostatic, antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV, antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic, antiparkinsonian and antilipaemic activities, and can be used in gene therapy. NOVX proteins are useful for treating or preventing a pathology associated with a NOVX protein in humans and for treating a syndrome associated with the human disease. NOVX nucleic acids, proteins and antibodies can be used in the treatment and diagnosis of cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis valve disease, tuberous sclerosis, scleroderma, obesity, transplantation disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis, infectious disease, anorexia, cancer-associated cachexia, cancer, Alzheimer's disease, Parkinson's disease, immune disorders, haematopoietic disorders, dyslipidaemias, and metabolic syndrome X. ACC62346 to ACC62465 represent PCR primers and probes for human NOVX sequences, which are used in examples from the present invention. ABR54277 represents a human trypsinogen protein given in comparison with the human NOV35b protein in the exemplification of the present invention

Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match	68.4%;	Score 1370;	DB 7;	Length 1650;
-------------	--------	-------------	-------	--------------

Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

QY	134	GCTCTTTCACCATGCGCTGGATCACTTCCTTTGAATGCAGAAAGCTTGCTGGCCAAAGATG	193
Db	11	GCTCTTTCACCATGCGCTGGATCACTTCCTTTGAATGCAGAAAGCTTGCTGGCCAAAGATG	70
QY	194	TGGGAATTGTGCCCCCTTGAGATCTATTCTTCTCAATATGTTGATCAAGCAGAGTTGG	253
Db	71	TTGGGATTGTGCCCCCTTGAGATCTATTCTTCTCAATATGTTGATCAAGCAGAGTTGG	130
QY	254	AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCTTGAGCCAGGCCAAGATGG	313
Db	131	AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCTTGAGCCAGGCCAAGATGG	190
QY	314	GCTTCTGCACAGATAGAGAAGATATTACTCTCTTTCATGACTGTGGTTCAGAATCTTA	373
Db	191	GCTTCTGCACAGATAGAGAAGATATTACTCTCTTTCATGACTGTGGTTCAGAATCTTA	250
QY	374	TGGAGAGAAATTAACCTTTCCTATGATTCGATTGGGCGGCTGGAAGTTGGAACAGACAA	433
Db	251	TGGAGAGAAATTAACCTTTCCTATGATTCGATTGGGCGGCTGGAAGTTGGAACAGACAA	310
QY	434	TCATCGACAAATCAAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA	493
Db	311	TCATCGACAAATCAAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA	370

QY	494	ATACAGATATAGAAGGAATGCACACAACCTAA	TGCATGCTAT	TGAGGACACAGCTGCTGTCT	553
Db	371	ATACAGATATAGAAGGAATGCACACAACCTAA	TGCTAT	TGAGGACACAGCTGCTGTCT	430
QY	554	TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGAT	-----	-----	591
Db	431	TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGAT	TGACCGGTATGCCCTG	GTAGTTG	490
QY	592	-----	-----	-----	591
Db	491	CAGAGATATTTGCTGATATGCCACAGGAATGCTAGACCTACAGGTGAGTTGGAGCAG	550		
QY	592	-----	-----	-----	592
Db	551	TAGCTCTGCTAATTTGGGCCAAATGCTCTTAAATTTTGAACGAGGGCTTCGTGGACAC	610		
QY	608	ATATGCAACATGCTTATGATTTTTCACAGCCCTGATATGCTATCTGAATATCTTATAGTAG	667		
Db	611	ATATGCAACATGCTTATGATTTTTCACAGCCCTGATATGCTATCTGAATATCTTATAGTAG	670		
QY	668	ATGGAATACTCTCCATACAGTGTCTACCTCAGTGTGATTAAGCCGCTGCTACTCTGTCTACT	727		
Db	671	ATGGAATACTCTCCATACAGTGTCTACCTCAGTGTGATTAAGCCGCTGCTACTCTGTCTACT	730		
QY	728	GCAAAAGATCCATGCGCCAGTGGCAGAAAGAGGGAATGATAAGATTTTACCTTGATG	787		
Db	731	GCAAAAGATCCATGCGCCAGTGGCAGAAAGAGGGAATGATAAGATTTTACCTTGATG	790		
QY	788	ATTTGGCTTCATGATCTTTTCACTCACCATATTTGTAACTGGTTCAGAAATCTCTAGCTC	847		
Db	791	ATTTGGCTTCATGATCTTTTCACTCACCATATTTGTAACTGGTTCAGAAATCTCTAGCTC	850		
QY	848	GGATGTGCTGAATGACTTCTTAAATGACCAGAAATAGATTAATAATAGTATCTATAGTG	907		
Db	851	GGATGTGCTGAATGACTTCTTAAATGACCAGAAATAGATTAATAATAGTATCTATAGTG	910		
QY	908	GCCGTGAAGCCTTTGGGATGTTAAATAGAGAACCCTACTTTGATAGATGTGAGA	967		
Db	911	GCCGTGAAGCCTTTGGGATGTTAAATAGAGAACCCTACTTTGATAGATGTGAGA	970		
QY	968	AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCAGTTCAGAAACAAAGGCATCTTACTTG	1027		
Db	971	AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCAGTTCAGAAACAAAGGCATCTTACTTG	1030		
QY	1028	TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTGCATCTGTTTC	1087		
Db	1031	TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTGCATCTGTTTC	1090		
QY	1088	TAGCAGTACTCACCCTGACCAATTAGCAGGGAAGAAATTGAGTGTCTTCTATGTT	1147		
Db	1091	TAGCAGTACTCACCCTGACCAATTAGCAGGGAAGAAATTGAGTGTCTTCTATGTT	1150		
QY	1148	CTGGTTGGCTGCCACTCTGTGTACTCTCTTAAAGTACACAAGATGCTACACGGGGCTCG	1207		
Db	1151	CTGGTTGGCTGCCACTCTGTGTACTCTCTTAAAGTACACAAGATGCTACACGGGGCTCG	1210		
QY	1208	CTCTTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTG	1267		
Db	1211	CTCTTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTG	1270		
QY	1268	GTGTGGCACAGATGTCCTTGGCTGAAAAACATGAAGCTCAGAGAGACACCCATCATTTGG	1327		
Db	1271	GTGTGGCACAGATGTCCTTGGCTGAAAAACATGAAGCTCAGAGAGACACCCATCATTTGG	1330		
QY	1328	TCACTATATTTCCCAAGGTTCAATAGATTCACTCTTGAAGAACTGGTACTTAGTTA	1387		
Db	1331	TCACTATATTTCCCAAGGTTCAATAGATTCACTCTTGAAGAACTGGTACTTAGTTA	1390		
QY	1388	GGGTGATGAAAAAGCACAGAAAGAACTTACGCTCGGCGTCCCACTCCAATGATGACACTT	1447		
Db	1391	GGGTGATGAAAAAGCACAGAAAGAACTTACGCTCGGCGTCCCACTCCAATGATGACACTT	1450		

QY 1448 TGGATGAAGAGTAGGACTTGTGCATTCAAACTAGCAACTGACATATTCCAAGCCCTG 1507
|||
Db 1451 TGGATGAAGAGTAGGACTTGTGCATTCAAACTAGCAACTGACATATTCCAAGCCCTG 1510
QY 1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTGATTAGTA 1567
|||
Db 1511 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTGATTAGTA 1570
QY 1568 ATGGGGAACATTAAAGTACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT 1627
|||
Db 1571 ATGGGGAACATTAAAGTACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT 1630
QY 1628 GGGGGTATGGGAACAGTTGG 1647
|||
Db 1631 GGGGGTATGGGAACAGTTGG 1650

RESULT 12

ACC62332
ID ACC62332 standard; cDNA; 1650 BP.

XX AC ACC62332;

XX DT 23-JUN-2003 (first entry)

XX DE Human NOV41g encoding cDNA SEQ ID NO:193.

XX Human; NOX; antiatherosclerotic; hypotensive; cardiac; dermatological;
KW anorectic; immunosuppressive; cytoslatic; antidiabetic; antiinfertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberosclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X; gene; ss.

XX OS Homo sapiens.

XX PN WO2003023001-A2.

XX PD 20-MAR-2003.

XX PF 09-SEP-2002; 2002WO-US028538.

XX 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 26-SEP-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 03-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.

XX (CURA-) CURAGEN CORP.
PA Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
XX Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eissen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Glot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zerhusen BD, Zhong M;
XX WPI; 2003-313241/30.
DR P-PSDB; ABR54263.

XX Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.

PS Claim 20; Page 243; 460pp; English.

XX The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiac,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytoslatic,
CC antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipemic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberosclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders.
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match 68.4%; Score 1370; DB 7; Length 1650;
Best Local Similarity 92.0%; Pred. No. 4.1e-259;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

QY 134 GCTCTTTCACCATGCGCTGGATCACTTCCTTGAATGACAGAGCTTGCGCCAAAGATG 193
|||
Db 11 GCTCTTTCACCATGCGCTGGATCACTTCCTTGAATGACAGAGCTTGCGCCAAAGATG 70
QY 194 TGGGAATTGTTGCCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCAGATTGG 253
|||
Db 71 TTGGATTGTTGCCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCAGATTGG 130
QY 254 AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG 313
|||
Db 131 AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG 190
QY 314 GCTTTCGACAGATAGAGAAGATATTAACTCTTTGCATGAGCTGTGTTCAAGATCTTA 373
|||
Db 191 GCTTTCGACAGATAGAGAAGATATTAACTCTTTGCATGAGCTGTGTTCAAGATCTTA 250
QY 374 TGGAGAGAAATTAACCTTCTTATGATTGATTTGGCGGCTGGAAGTTGGAACAGACAA 433
|||
Db 251 TGGAGAGAAATTAACCTTCTTATGATTGATTTGGCGGCTGGAAGTTGGAACAGACAA 310

QY

434

TCATCGACAATCAAAGCTGTGAGACTAATTGTGACGCTGTTGAAGAGCTGGGA

493

Db

311

TCATCGACAATCAAAGCTGTGAGACTAATTGTGACGCTGTTGAAGAGCTGGGA

370

QY

494

ATACAGATATAGAAGGAATCGACAACAATAATGCATGCTATGGAGCACAGCTGCTCT

553

Db

371

ATACAGATATAGAAGGAATCGACAACAATAATGCATGCTATGGAGCACAGCTGCTCT

430

QY

554

TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGAT-----

591

Db

431

TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGATGACGGTATGCCCTGTAATTG

490

QY

592

591

Db

491

CAGGAGATATTGCTGTAATGCCACAGGAATGCTAGACCTACAGGTGAGTTGAGACAG

550

QY

592

-----GGGCTTCGTGGGACAC

607

Db

551

TAGCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGCTTCGTGGACAC

610

QY

608

ATATGCAACATGCCCTATGATTTTTCAAAGCCTGATATGCTATCTGAATATCCTATAGTAG

667

Db

611

ATATGCAACATGCCCTATGATTTTTCAAAGCCTGATATGCTATCTGAATATCCTATAGTAG

670

QY

668

ATGGA AAACTCTCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT

727

Db

671

ATGGA AAACTCTCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT

730

QY

728

GCAAAAAGATCCATGCCCGCAGTGGCAGAAAGAGGGAATGATAAGATTTTACCCTTGAATG

787

Db

731

GCAAAAAGATCCATGCCCGCAGTGGCAGAAAGAGGGAATGATAAGATTTTACCCTTGAATG

790

QY

788

ATTTGGCTTCATGATCTTTCACCTCACCATATTTGTAACCTGGTTCAGAAATCTCTAGCTC

847

Db

791

ATTTGGCTTCATGATCTTTCACCTCACCATATTTGTAACCTGGTTCAGAAATCTCTAGCTC

850

QY

848

GGATGTTGCTGAATGACTTCCTTAATGACCAGAAATAGAGATAAAATAGTATCTATAGTG

907

Db

851

GGATGTTGCTGAATGACTTCCTTAATGACCAGAAATAGAGATAAAATAGTATCTATAGTG

910

QY

908

GCCTGGAAGCCTTTGGGGAGTTAAATTAGAAGACACCTACTTGTATAGAGATGGAGA

967

Db

911

GCCTGGAAGCCTTTGGGGAGTTAAATTAGAAGACACCTACTTGTATAGAGATGGAGA

970

QY

968

AGGCATTTATGAAGGCTAGCTCTGAACCTCTCAGTCAGAAAACAAAGGCATCTTACTTG

1027

Db

971

AGGCATTTATGAAGGCTAGCTCTGAACCTCTCAGTCAGAAAACAAAGGCATCTTACTTG

1030

QY

1028

TATCAAAATCAAATGGAATATGTACACATCTTCAGTATATGCTTCCCTTGCACTCTGTTTC

1087

Db

1031

TATCAAAATCAAATGGAATATGTACACATCTTCAGTATATGCTTCCCTTGCACTCTGTTTC

1090

QY

1088

TAGCACAGTACTCACCCTCAGCAATTAGCAGGGAAGAAATGGAGTGTTCCTTATGCTT

1147

Db

1091

TAGCACAGTACTCACCCTCAGCAATTAGCAGGGAAGAAATGGAGTGTTCCTTATGCTT

1150

QY

1148

CTGGTTTGGCTGCCACTCTGTACTCTCTTAAGTCAACAAGATGCTACACCGGGGTCTG

1207

Db

1151

CTGGTTTGGCTGCCACTCTGTACTCTCTTAAGTCAACAAGATGCTACACCGGGGTCTG

1210

QY

1208

CTCTTGATAAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTG

1267

Db

1211

CTCTTGATAAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTG

1270

QY

1268

GTTGGCACCAAGATGTCCTCGCTGA AAAACATGAAAGCTCAGAGAGGACACCCATCATTTGG

1327

Db

1271

GTTGGCACCAAGATGTCCTCGCTGA AAAACATGAAAGCTCAGAGAGGACACCCATCATTTGG

1330

QY

1328

TCAAATAATATCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTATGTTA

1387

Db

1331

TCAAATAATATCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTATGTTA

1390

QY

1388

GGGTGGATGAAAAGCACAGAAGAATTAACGCTCGGCGTCCCACTCCAATGATGACACTT

1447

Db

1391

GGGTGGATGAAAAGCACAGAAAGAACTTACGCTCGGCGTCCCACTCCAATGATGACACTT

1450

QY

1448

TGGATGAAGAGTAGACTTGTGCATTCAACATAGCAACTGAGCATATTTCCAAGCCCTG

1507

Db

1451

TGGATGAAGAGTAGACTTGTGCATTCAACATAGCAACTGAGCATATTTCCAAGCCCTG

1510

QY

1508

CCAAGAAAGTACCAGACTTCCTGCCACAGCAGCAGAACTGAAAGCAGCTGTCAATTAGTA

1567

Db

1511

CCAAGAAAGTACCAGACTTCCTGCCACAGCAGCAGAACTGAAAGCAGCTGTCAATTAGTA

1570

QY

1568

ATGGGGAACATTAAAGATACTGTGAGGTGCAAGACTTCAGGGTGGGTGGGCATGGGGT

1627

Db

1571

ATGGGGAACATTAAAGATACTGTGAGGTGCAAGACTTCAGGGTGGGTGGGCATGGGGT

1630

QY

1628

GGGGGTATGGGAACAGTTGG

1647

Db

1631

GGGGGTATGGGAACAGTTGG

1650

RESULT 13

ADE38394

ID ADE38394 standard; DNA; 1650 BP.

XX ADE38394;

AC

AC

DT 29-JAN-2004 (first entry)

XX

DE Human protein 9389 gene sequence.

XX

KW tumourigenic disorder; angiogenic disorder; aberrant gene expression;

KW aberrant protein activity; cytostatic; antihyroid; antidiabetic;

KW ophthalmological; cancer; breast cancer; colon cancer; lung cancer;

KW prostatic cancer; Grave's disease; diabetic retinopathy; gene; ds;

KW protein 9389.

XX Homo sapiens.

OS

XX

FH Key

FT CDS

FT

FT

XX

PN WO2003065006-A2.

XX

PD 07-AUG-2003.

XX

PF 30-JAN-2003; 2003WO-US002588.

XX

Location/Qualifiers

22. .1584

/*tag= a

/product= "Human protein 9389"

31-JAN-2002; 2002US-0353600P.

PR 15-MAR-2002; 2002US-0364517P.

PR 09-APR-2002; 2002US-0371075P.

PR 10-APR-2002; 2002US-0371507P.

PR 16-APR-2002; 2002US-0372984P.

PR 19-APR-2002; 2002US-0374194P.

PR 24-MAY-2002; 2002US-0382995P.

PR 31-MAY-2002; 2002US-0385023P.

PR 14-JUN-2002; 2002US-0388853P.

PR 17-JUN-2002; 2002US-0389395P.

PR 25-JUN-2002; 2002US-0391324P.

PR 15-JUL-2002; 2002US-0395944P.

PR 22-JUL-2002; 2002US-0397726P.

PR 13-AUG-2002; 2002US-0403046P.

PR 22-AUG-2002; 2002US-0405155P.

PR 27-AUG-2002; 2002US-0406361P.

PR 25-OCT-2002; 2002US-0421195P.

PR 12-NOV-2002; 2002US-0425456P.

PR 19-NOV-2002; 2002US-0427626P.

PR 10-DEC-2002; 2002US-0432122P.

XX

PA (MILL-) MILLENNIUM PHARM INC.

XX

PI Hunter JÜ, Macbeth KJ, Tsai F, Lesoon A, Lightcap ES;

PI Williamson MW, Rudolph-Owen LA;
XX WPI: 2003-646176/61.
DR P-PSDB; ADE38395.
XX
PT Treating subject having tumorigenic disorder or angiogenic disorder
PT caused by aberrant polypeptide e.g., N-Formylpeptide receptor or nucleic
PT acid, by administering a modulator.
XX
XX Disclosure; SEQ ID NO 55; 454bp; English.
XX This invention relates to a novel method of treating a human subject
CC having a tumorigenic disorder or angiogenic disorder, caused by aberrant
CC gene expression or activity of an isolated protein, by administering a
CC modulator. The modulator may have cytostatic, antithyroid, antidiabetic
CC or ophthalmological activity. The method is useful for treating a subject
CC having a tumorigenic or angiogenic disorder, in particular for treating
CC cancer (for example breast cancer, colon cancer, lung cancer or prostatic
CC cancer) and, for example, Grave's disease and diabetic retinopathy. The
CC present sequence is a DNA sequence which encodes the novel isolated human
CC protein 9389 of the invention.
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match 68.4%; Score 1370; DB 9; Length 1650;
Best Local Similarity 92.0%; Pred. No. 4.1e-259;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

QY 134 GCTCTTTCACCATGCGCTGGATCACTTCTTGAATGCAGAAAGCTTGCGCCAAAAGATG 193
DB 11 GCTCTTCACCATGCGCTGGATCACTTCTTGAATGCAGAAAGCTTGCGCCAAAAGATG 70
QY 194 TGGGAATTGTTGCCCTTGAGATCTATTTCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
DB 71 TTGGGATTGTTGCCCTTGAGATCTATTTCTTCTCAATATGTTGATCAAGCAGAGTTGG 130
QY 254 AAAAATATGATGCTGTAGATGCTGGAAGTATACCATTTGGCTTGGCCAGGCCAAGATGG 313
DB 131 AAAAATATGATGCTGTAGATGCTGGAAGTATACCATTTGGCTTGGCCAGGCCAAGATGG 190
QY 314 GCTTCTGCACAGATAGAGAATATTAACCTCTCTTGCATGACTGTGTTCAAGATCTTA 373
DB 191 GCTTCTGCACAGATAGAGAATATTAACCTCTCTTGCATGACTGTGTTCAAGATCTTA 250
QY 374 TGGAGAGAAATAACCTTCCATGATTGCAATGGCGCGCTGGAAGTTGGAACAGAGACAA 433
DB 251 TGGAGAGAAATAACCTTCCATGATTGCAATGGCGCGCTGGAAGTTGGAACAGAGACAA 310
QY 434 TCATCGACAAATCAAAAGCTCTGGAAGACTAATTTGATGCAGCTGTTGAAGAGTCTGGGA 493
DB 311 TCATCGACAAATCAAAAGCTCTGGAAGACTAATTTGATGCAGCTGTTGAAGAGTCTGGGA 370
QY 494 ATACAGATATAGAAAGGAATGCAGACAACTAATGCATGCTATGAGGCAACAGCTGCTGCT 553
DB 371 ATACAGATATAGAAAGGAATGCAGACAACTAATGCATGCTATGAGGCAACAGCTGCTGCT 430
QY 554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT----- 591
DB 431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGAACGCTATGCCCTGTAAGTTG 490
QY 592 ----- 591
DB 491 CAGGAGATATTGCTGTATATGCCACAAGAAATGCTAGACCTACAGGTGAGGTGAGCAG 550
QY 592 -----GGGCTTCGTGGAGAC 607
DB 551 TAGCTCTGCTAATTGGGCCAAATGCTCTTAAATTTTGAACGAGGGCTTCGTGGAGAC 610
QY 608 ATATGCAACATGCTATGATTTTAAACAGCCTGATATGCTATGATATCCTAATAGTAG 667
DB 611 ATATGCAACATGCTATGATTTTAAACAGCCTGATATGCTATGATATCCTAATAGTAG 670
QY 668 ATGGAATACTCTCCATACAGTGTCACTCAGTGCATTAGACCGCTGCTACTCTGTCTACT 727

DB 671 ATGGAAACTCTCCATAGAGTGTACTACCTGACGTGCAATTAACCGCTGCTATTCGTCTACT 730
QY 728 GCAAAAAAGATCCATGCCCCAGTGCGAGAAAGGGAATGATAAGATTTTACCTGGAATG 787
DB 731 GCAAAAAAGATCCATGCCCCAGTGCGAGAAAGGGAATGATAAGATTTTACCTGGAATG 790
QY 788 ATTTGGCTTCATGATCTTTCACTCACTCACTAATTTGTAACTGCTCAGAAATCTTAGCTC 847
DB 791 ATTTGGCTTCATGATCTTTCACTCACTCACTAATTTGTAACTGCTCAGAAATCTTAGCTC 850
QY 848 GGATGTTGCTGAATGACTTCTCTTAATGACCAAGATAGAGATAAATAATGATCTATAGTG 907
DB 851 GGATGTTGCTGAATGACTTCTCTTAATGACCAAGATAGAGATAAATAATGATCTATAGTG 910
QY 908 GCCTGGAAGCCTTTGGGATGTTAATTAGAAACAACCTACTTTGATAGAGATGAGAGA 967
DB 911 GCCTGGAAGCCTTTGGGATGTTAATTAGAAACAACCTACTTTGATAGAGATGAGAGA 970
QY 968 AGGCATTATGAAGGCTAGCTCTGAACTCTTCACTCAGTCAAGAAACAAAGGATCTTACTTG 1027
DB 971 AGGCATTATGAAGGCTAGCTCTGAACTCTTCACTCAGTCAAGAAACAAAGGATCTTACTTG 1030
QY 1028 TATCAATATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTTC 1087
DB 1031 TATCAATATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTTC 1090
QY 1088 TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAAATGGAAGTGTTCCTAATGTT 1147
DB 1091 TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAAATGGAAGTGTTCCTAATGTT 1150
QY 1148 CTGGTTGGCTGCCACTCTGTACTCTCTTAAAGTACACAAAGATGCTACACCGGGTCTG 1207
DB 1151 CTGGTTGGCTGCCACTCTGTACTCTCTTAAAGTACACAAAGATGCTACACCGGGTCTG 1210
QY 1208 CTCTTGATAAATAAACAAGCAAGTTATGATCTTAAATCAAGGCTTGATTCAGAACTG 1267
DB 1211 CTCTTGATAAATAAACAAGCAAGTTATGATCTTAAATCAAGGCTTGATTCAGAACTG 1270
QY 1268 GTGTGGCACCAAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGCAACCCATCATTTGG 1327
DB 1271 GTGTGGCACCAAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGCAACCCATCATTTGG 1330
QY 1328 TCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTAATTAGTTA 1387
DB 1331 TCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTAATTAGTTA 1390
QY 1388 GGGTGATGAAAGACAGAGAAGCTTACGCTCGGCGTCCCACTCCAATGATGACACTT 1447
DB 1391 GGGTGATGAAAGACAGAGAAGCTTACGCTCGGCGTCCCACTCCAATGATGACACTT 1450
QY 1448 TCGATGAAGGAGTAGGACTTGTGATTCAAACTAGCAACTGACATATTTCCAAGCCCTG 1507
DB 1451 TCGATGAAGGAGTAGGACTTGTGATTCAAACTAGCAACTGACATATTTCCAAGCCCTG 1510
QY 1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGACAGAACTGAAAGCAAGCTGTCAATAGTA 1567
DB 1511 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGACAGAACTGAAAGCAAGCTGTCAATAGTA 1570
QY 1568 ATGGGAACTTAAGATACTCTGTAGAGGTGCAAGACTTCAAGGTGGGGGTGGGATGGGGT 1627
DB 1571 ATGGGAACTTAAGATACTCTGTAGAGGTGCAAGACTTCAAGGTGGGGGTGGGATGGGGT 1630
QY 1628 GGGGCTATGGGAACAGTTGG 1647
DB 1631 GGGGCTATGGGAACAGTTGG 1650

RESULT 14
AAH34834
ID AAH34834 standard; cDNA; 3008 BP.
XX
AC AAH34834;

XX 03-SEP-2001 (first entry)
DT
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1916.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 5; 88.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR P-PSDB; AAG75429.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 3427-3428; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG7778 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 3008 BP; 944 A; 485 C; 638 G; 928 T; 0 U; 13 Other;
Query Match 66.1%; Score 1322.4; DB 4; Length 3008;
Best Local Similarity 97.7%; Pred. No. 9.5e-250;
Matches 1373; Conservative 9; Mismatches 20; Indels 4; Gaps 4;

Db 460 CAGAAATCTCTAGCTCGATGTGCTGTAATGACTTCCCTTAATGACCAGATAGAGATAA 519
Qy 892 AATAGTATCTATAGTGGCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTT 951
Db 520 AATAGTATCTATAGTGGCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTT 579
Qy 952 GATTAGAGATGTGGAAGGCATTATGAAGGCTAGCTGAACTCTTCAGTCAGAAAAA 1011
Db 580 GATTAGAGATGTGGAAGGCATTATGAAGGCTAGCTGAACTCTTCAGTCAGAAAAA 639
Qy 1012 AAGGCATCTTTACTTGTATCAAAATCAAAATGAAATGTACACATCTTCAGTATATG 1071
Db 640 AAGGCATCTTTACTTGTATCAAAATCAAAATGAAATGTACACATCTTCAGTATATG 699
Qy 1072 TCCCTTGATCTGTTCTAGACAGTACTCACCTCAGCAATTAGCAGGAAGAATTGGA 1131
Db 700 TCCCTTGATCTGTTCTAGACAGTACTCACCTCAGCAATTAGCAGGAAGAATTGGA 759
Qy 1132 GTGTTTCTTATGTTCTGGTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAGAT 1191
Db 760 GTGTTTCTTATGTTCTGGTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAGAT 819
Qy 1192 GCTACACCGGGGTCTGCTTGTATAAATTAACAGCAAGTTATGTATCTTAAAT-CAAG 1250
Db 820 GCTACACCGGGGTCTGCTTGTATAAATTAACAGCAAGTTATGTATCTTAAATCAAG 879
Qy 1251 GCTTGATT-CAAGAACTGGTGGCCACAGATGTCCTGCTGAAACATGAAGCT-CA 1308
Db 880 GCTTGATTCCAAAGAACTGGTGGCCACAGATGTCCTGCTGAAACATGAAGCTCCA 939
Qy 1309 GAGGACACCCATCATTTGGTCAACTATATTTCCCGAGGTTCAATAGATTCACTCTTGA 1368
Db 940 GAGGACACCCATCATTTGGTCAACTATATTTCCCGAGGTTCAATAGATTCACTCTTGA 999
Qy 1369 GGAACGTGTTACTTATGTTAGGTGATGAAAGCACAGAAAGCTTACGCTGGGCTCCC 1428
Db 1000 GGAACGTGTTACTTATGTTAGGTGATGAAAGCACAGAAAGCTTACGCTGGGCTCCC 1059
Qy 1429 ACTCCAATGATGACACCTTTGGATGAAGAGTAGACTTGTCAATCAACATAGCAACT 1488
Db 1060 ACTCCAATGRTGACACTTTGGATGAAGAGTAGACTTGTCAATCAACATAGCAACT 1119
Qy 1489 GAGCATATTTCCAAGCCCTGCCAAGAAAGTACCAGAATCCCTGCCACAGCAGAAACCT 1548
Db 1120 GAGCATATTTCCAAGCCCTGCCAAGAAAGTACCAGAATCCCTGCCACAGCAGAAACCT 1179
Qy 1549 GAAGCAGCTGTCAATAGTAAATGGGAAACATTAAGATACCTGTGAGGTGCAAGACTT 1608
Db 1180 GAAGCAGCTGTCAATAGTAAATGGGAAACATTAAGATACCTGTGAGGTGCAAGACTT 1239
Qy 1609 GGTGGGTGGGATGGGTGGGATGGGAAACAGTTGGAGAAATGGATATCTGGGAT 1668
Db 1240 GGTGGGTGGGATGGGTGGGATGGGAAACAGTTGGAGAAATGGATATCTGGGAT 1299
Qy 1669 AATTTAAAGGATTACATGTTATGTAAATTTTATGTGACTGACATGAGCCTGATGAC 1728
Db 1300 AATTTAAAGGATTACATGTTATGTAAATTTTATGTGACTGACATGAGCCTGATGAC 1359
Qy 1729 TATCGTACTTGGGAAAGTCTTTGCTCTATTTGTCGACATGCTTCTGTGTGCTCT 1788
Db 1360 TATCGTACTTGGGAAAGTCTTTGCTCTATTTGTCGACATGCTTCTGTGTGCTCT 1419
Qy 1789 GGCCAATGCCAATGTACTCGAATGATGTTAAGGGCTGTGTAACCTTCATACCTCTTTG 1848
Db 1420 GGCCAATGCCAATGTACTCGAATGATGTTAAGGGCTGTGTAACCTTCATACCTCTTTG 1479
Qy 1849 GCCATTTGTATGATGATGTTTGGTTTAAACATGTTAATGAATGTGTAATCTGTTCT 1908
Db 1480 GCCATTTGTATGATGATGTTTGGTTTAAACATGTTAATGAATGTGTAATCTGTTCT 1539
Qy 1909 CAGAAAGACAGAGGTACTAATCTCCAATTAAAAAATTTTAAACATGTAAAAA 1968

Db 1540 CAGAAAGMAAGCAGAGGTACTAATCTCCAAATTAAAAATTTTAACATGTAGAAATTTTG 1599

QY 1969 AAAAAAAAAAAAAAAAAAAAAA 1994

Db 1600 TACTTTGAACACACAGATTACAGAAA 1625

RESULT 15

ACC62337

ID ACC62337 standard; cDNA; 1564 BP.

XX

AC ACC62337;

XX

DT 23-JUN-2003 (first entry)

XX

DE Human NOV41L encoding cDNA SEQ ID NO:203.

XX

KW Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;

KW anorectic; immunosuppressive; cyostatic; antidiabetic; antinfertility;

KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;

KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeimic;

KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;

KW congenital heart defect; aortic stenosis; valve disease; transplantation;

KW tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;

KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;

KW fertility; haemophilia; hypercoagulation; graft versus host disease;

KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;

KW Crohn's disease; multiple sclerosis; infectious disease; cancer;

KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;

KW immune disorder; haematopoietic disorder; dyslipidaemia;

KW metabolic syndrome X; gene; ss.

XX

OS Homo sapiens.

XX

PN WO2003023001-A2.

XX

20-MAR-2003.

PD

XX

PF 09-SEP-2002; 2002WO-US028538.

XX

PR 07-SEP-2001; 2001US-0318120P.

PR 07-SEP-2001; 2001US-0318184P.

PR 10-SEP-2001; 2001US-0318430P.

PR 17-SEP-2001; 2001US-0322636P.

PR 17-SEP-2001; 2001US-0322781P.

PR 17-SEP-2001; 2001US-0322816P.

PR 17-SEP-2001; 2001US-0322817P.

PR 19-SEP-2001; 2001US-0323519P.

PR 20-SEP-2001; 2001US-0323631P.

PR 20-SEP-2001; 2001US-0323636P.

PR 25-SEP-2001; 2001US-0324969P.

PR 25-SEP-2001; 2001US-0325091P.

PR 26-SEP-2001; 2001US-0324990P.

PR 14-DEC-2001; 2001US-0341144P.

PR 26-FEB-2002; 2002US-0359599P.

PR 05-MAR-2002; 2002US-0361663P.

PR 03-MAY-2002; 2002US-0377908P.

PR 17-MAY-2002; 2002US-0381483P.

PR 29-MAY-2002; 2002US-0383863P.

PR 02-JUL-2002; 2002US-0393332P.

PR 17-JUL-2002; 2002US-0396412P.

PR 13-AUG-2002; 2002US-0403517P.

PR 06-SEP-2002; 2002US-00236417.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;

PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;

PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;

PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;

PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;

PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patcurajan M;

PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;

PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;

PI Zernusen BD, Zhong M;

XX

DR WPI; 2003-313241/30.

DR P-PSDB; ABR54268.

XX

PT Novel human proteins and nucleic acid encoding the proteins, useful for

PT diagnosis, treatment and prevention of disorders involving the human

PT protein or nucleic acid e.g. cardiac and neurological disorders.

XX

PS Claim 20; Page 245-246; 460pp; English.

CC The present invention describes isolated human NOVX proteins, where X is

CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in

CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,

CC hypotensive, dermatological, anorectic, immunosuppressive, cyostatic,

CC antidiabetic, antinfertility, haemostatic, antiinflammatory, anti-HIV,

CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,

CC antiparkinsonian and antilipaeimic activities, and can be used in gene

CC therapy. NOVX proteins are useful for treating or preventing a pathology

CC associated with a NOVX protein in humans and for treating a syndrome

CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,

CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,

CC valve disease, tuberous sclerosis, scleroderma, obesity, metabolic

CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic

CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,

CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host

CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,

CC infectious disease, anorexia, cancer-associated cachexia, cancer,

CC Alzheimer's disease, Parkinson's disease, immune disorders,

CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.

CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX

CC sequences, which are used in examples from the present invention.

CC ABR54277 represents a human trypsinogen protein given in comparison with

CC the human NOV35b protein in the exemplification of the present invention

XX

SQ Sequence 1564 BP; 464 A; 304 C; 362 G; 434 T; 0 U; 0 Other;

Query Match 65.0%; Score 1302; DB 7; Length 1564;

Best Local Similarity 91.9%; Pred. No. 8.6e-246;

Matches 1438; Conservative 0; Mismatches 0; Indels 126; Gaps 1;

QY 144 CATGCCCTGGATCACTTCCTTGAATGCAGAACTTGCTGGCCAAAAGATGTGGAATTGT 203

Db 1 CATGCCCTGGATCACTTCCTTGAATGCAGAACTTGCTGGCCAAAAGATGTGGAATTGT 60

QY 204 TGCCCTTGAGATCTAATTTCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGA 263

Db 61 TGCCCTTGAGATCTAATTTCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGA 120

QY 264 TGGTGTAGATGCTGGAAGATATACCATTTGGCTGGGCCAGGCCAAGATGGGCTTGACAC 323

Db 121 TGGTGTAGATGCTGGAAGATATACCATTTGGCTGGGCCAGGCCAAGATGGGCTTGACAC 180

QY 324 AGATAGAGAAGATATTAACTCTCTTGCATGACTGTGTTGAGAACTTTATGAGAGAAAA 383

Db 181 AGATAGAGAAGATATTAACTCTCTTGCATGACTGTGTTGAGAACTTTATGAGAGAAAA 240

QY 384 TAACCTTTCTATGATTGCAATGGGGCGGCTGGAAGTTGGAACAGACAATCATCGACAA 443

Db 241 TAACCTTTCTATGATTGCAATGGGGCGGCTGGAAGTTGGAACAGACAATCATCGACAA 300

QY 444 ATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGCTCGGAATACAGATAT 503

Db 301 ATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGCTCGGAATACAGATAT 360

QY 504 AGAAGGAATGCACACAACTATATGCATGCTATGAGGACACAGCTGCTTCAATGCTGT 563

Db 361 AGAAGGAATGCACACAACTATATGCATGCTATGAGGACACAGCTGCTTCAATGCTGT 420

QY 564 TAACTGGATTGAGTCCAGCTCTTGGGAT----- 591

Db	421	TTAACTGGATTGAGTCCAGCTCTTGGGATGACGGTATGCGCTGTAGTTGCAGGAGATAT	480
QY	592	-----	591
Db	481	TGCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGAGCAGTAGCTGTGCT	540
QY	592	-----GGGCTTCGTGGGACACATATGCAACA	617
Db	541	AATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACA	600
QY	618	TGCTTATGATTTTTCACAGCCCTGATATGCTATCTGAATATCCTATAGATGGAATACT	677
Db	601	TGCTTATGATTTTTCACAGCCCTGATATGCTATCTGAATATCCTATAGATGGAATACT	660
QY	678	CTCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTACTGTCTACTGCAAAAAAGAT	737
Db	661	CTCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTACTGTCTACTGCAAAAAAGAT	720
QY	738	CCATGCCCCAGTGGCAGAAAGGGAATGATAAAGATTTTAACTTGAATGATTTTGCTT	797
Db	721	CCATGCCCCAGTGGCAGAAAGGGAATGATAAAGATTTTAACTTGAATGATTTTGCTT	780
QY	798	CATGATCTTTCACTCACCATATGTAACTGGTTCAGAAATCTCTAGCTCGAGTGTGCT	857
Db	781	CATGATCTTTCACTCACCATATGTAACTGGTTCAGAAATCTCTAGCTCGAGTGTGCT	840
QY	858	GAATGACTTCCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTGGCTGGAAGC	917
Db	841	GAATGACTTCCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTGGCTGGAAGC	900
QY	918	CTTTGGGGATGTTAAATTAGAGACACCTACTTTGATAGAGATGGAAGGCATTTAT	977
Db	901	CTTTGGGGATGTTAAATTAGAGACACCTACTTTGATAGAGATGGAAGGCATTTAT	960
QY	978	GAAAGCTAGCTCTGAACCTTTCACTCAGAAAACAAAAGCATCTTAACTTGATCAAAATCA	1037
Db	961	GAAAGCTAGCTCTGAACCTTTCACTCAGAAAACAAAAGCATCTTAACTTGATCAAAATCA	1020
QY	1038	AAATGGAATATGTACACACATCTTCAGTATATGTTCCCTTGCACTGTCTAGCACAGTA	1097
Db	1021	AAATGGAATATGTACACACATCTTCAGTATATGTTCCCTTGCACTGTCTAGCACAGTA	1080
QY	1098	CTCACTCAGCAATTAGCAGGGAAGAAATTGAGTGTCTTCTTATGGTCTGCTTGGC	1157
Db	1081	CTCACTCAGCAATTAGCAGGGAAGAAATTGAGTGTCTTCTTATGGTCTGCTTGGC	1140
QY	1158	TGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGCTGCTCTTGATTA	1217
Db	1141	TGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGCTGCTCTTGATTA	1200
QY	1218	AATTAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATCAAGAACTGCTGGCAC	1277
Db	1201	AATTAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATCAAGAACTGCTGGCAC	1260
QY	1278	AGATGTCTTCGCTGAAAACATGAACTCAGAGAGACACCCATCATTTGGTCAACTATAT	1337
Db	1261	AGATGTCTTCGCTGAAAACATGAACTCAGAGAGACACCCATCATTTGGTCAACTATAT	1320
QY	1338	TCCCCAGGGTTCAATAGATTCACTTTGAAGGAACGTGTTACTTAGTTAGGGTGATGA	1397
Db	1321	TCCCCAGGGTTCAATAGATTCACTTTGAAGGAACGTGTTACTTAGTTAGGGTGATGA	1380
QY	1398	AAAGCACAGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAG	1457
Db	1381	AAAGCACAGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAG	1440
QY	1458	AGTAGAATTGTGCATTCAAAACATAGCAACTGAGCATATTTCCAAGCCCTGCCAAGAAAGT	1517
Db	1441	AGTAGAATTGTGCATTCAAAACATAGCAACTGAGCATATTTCCAAGCCCTGCCAAGAAAGT	1500
QY	1518	ACCAAGACTCCCTGCCACAGCAGCAGAACTGAAAGCAGCTGTCAATTAGTAATGGGAAACA	1577
Db	1501	ACCAAGACTCCCTGCCACAGCAGCAGAACTGAAAGCAGCTGTCAATTAGTAATGGGAAACA	1560

QY	1578	TTAA	1581
Db	1561	TTAA	1564

Search completed: June 24, 2004, 07:33:29
Job time : 783 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2004, 06:29:36 ; Search time 171 Seconds
(without alignments)
6497.144 Million cell updates/sec

Title: US-10-622-516-1

Perfect score: 2002
Sequence: 1 cgcctccacgagcactctcgg.....aaaaaaaaaaaaaaaa 2002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2002	100.0	2002	4	US-09-819-993-1	Sequence 1, Appli
2	2002	100.0	2002	4	US-10-193-295-1	Sequence 1, Appli
3	794.6	39.7	1824	1	US-08-305-505-1	Sequence 1, Appli
4	476	23.8	28001	4	US-09-819-993-3	Sequence 3, Appli
5	476	23.8	28001	4	US-10-193-295-3	Sequence 3, Appli
6	436.4	21.8	615	4	US-09-370-838-245	Sequence 245, App
7	410	20.5	500	4	US-09-370-838-128	Sequence 128, App
8	401	20.0	506	4	US-09-401-064-187	Sequence 187, App
9	224	11.2	448	4	US-09-833-381-1645	Sequence 1645, Ap
10	210.4	10.5	472	4	US-09-833-381-1648	Sequence 1648, Ap
11	135.8	6.8	635	3	US-08-998-416-730	Sequence 730, App
12	110	5.5	307	4	US-08-998-416-744	Sequence 744, App
13	103.6	5.2	4775	3	US-09-306-595C-1	Sequence 1, Appli
14	103.6	5.2	4775	4	US-09-925-388-1	Sequence 1, Appli
15	86.6	4.3	305	4	US-09-313-294A-4313	Sequence 4313, Ap
16	74.6	3.7	385	4	US-09-833-381-1646	Sequence 1646, Ap
17	71.2	3.6	288	4	US-09-313-294A-5701	Sequence 5701, Ap
18	70.4	3.5	593	4	US-09-904-615-59	Sequence 59, Appl
19	70	3.5	293	4	US-09-313-294A-7134	Sequence 7134, Ap
20	67.2	3.4	232	4	US-09-833-381-745	Sequence 745, App
21	64.4	3.2	664	4	US-09-904-615-66	Sequence 66, Appl
22	64	3.2	282	4	US-09-621-976-18648	Sequence 18648, A
23	63.4	3.2	2550	6	5258287-23	Patent No. 5258287
24	62	3.1	6409	4	US-09-967-908A-1	Sequence 1, Appli
25	59.8	3.0	1474	3	US-08-821-994-64	Sequence 64, Appl
26	59.6	3.0	194	4	US-09-621-976-9596	Sequence 9596, Ap
27	59.6	3.0	3581	2	US-08-738-349-1	Sequence 1, Appli

28	59.2	3.0	1582	3	US-08-545-196B-10	Sequence 10, Appl
29	59.2	3.0	1582	3	US-08-545-196B-12	Sequence 12, Appl
30	59.2	3.0	2186	4	US-09-360-545-66	Sequence 66, Appl
31	59	2.9	1049	4	US-09-800-729-67	Sequence 67, Appl
32	58.8	2.9	1641	1	US-08-300-903A-8	Sequence 8, Appli
33	58.8	2.9	1641	4	US-08-988-197-8	Sequence 8, Appli
34	58.8	2.9	1736	3	US-09-182-816-22	Sequence 22, Appl
35	58.8	2.9	1736	3	US-09-182-816-24	Sequence 24, Appl
36	58.8	2.9	1736	3	US-09-471-528-22	Sequence 22, Appl
37	58.8	2.9	1736	3	US-09-471-528-24	Sequence 24, Appl
38	58.8	2.9	1736	3	US-09-634-530-22	Sequence 22, Appl
39	58.8	2.9	1736	3	US-09-634-530-24	Sequence 24, Appl
40	58.8	2.9	2246	4	US-09-363-708-3	Sequence 3, Appli
41	58.8	2.9	2246	4	US-09-083-587-3	Sequence 3, Appli
42	58.6	2.9	1074	3	US-09-248-335-67	Sequence 67, Appl
43	58.2	2.9	441	4	US-09-601-537-10	Sequence 10, Appl
44	58.2	2.9	1117	3	US-09-247-373B-33	Sequence 33, Appl
45	58.2	2.9	4121	4	US-09-601-537-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-09-819-993-1
; Sequence 1, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819, 993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Human
; US-09-819-993-1

Query Match	100.0%;	Score 2002;	DB 4;	Length 2002;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2002;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	CGCCTCCACGACTCTCGGAGTCCGGAGTCGGGTGGCGGCTATAAAGCTGGT	60	
Db	1	CGCCTCCACGACTCTCGGAGTCCGGAGTCGGGTGGCGGCTATAAAGCTGGT	60	
QY	61	AGCGAAGGGAGGCGCGCGGACTGTCTTCGTGCTCACTCCCTTCTGTCGCG	120	
Db	61	AGCGAAGGGAGGCGCGCGGACTGTCTTCGTGCTCACTCCCTTCTGTCGCG	120	
QY	121	CTCGGTACGCTTGCTCTTTCACCATGCTTGATCACTTCTTGAATGCAGAAGCTTGC	180	
Db	121	CTCGGTACGCTTGCTCTTTCACCATGCTTGATCACTTCTTGAATGCAGAAGCTTGC	180	
QY	181	TGGCCAAAAGATGTGGAAATGTTGCCCTTGAGATCTATTCTCTCAATATGTTGAT	240	
Db	181	TGGCCAAAAGATGTGGAAATGTTGCCCTTGAGATCTATTCTCTCAATATGTTGAT	240	
QY	241	CAAGCAGAGTTGAAAAATATGATGTGTAGATGCTGGAAGTATACCATTTGGCTGGGC	300	
Db	241	CAAGCAGAGTTGAAAAATATGATGTGTAGATGCTGGAAGTATACCATTTGGCTGGGC	300	
QY	301	CAGGCCAAGATGGGCTTCTGCACAGATAGAGAAGATTTAACTCTTTGATGACTGTG	360	
Db	301	CAGGCCAAGATGGGCTTCTGCACAGATAGAGAAGATTTAACTCTTTGATGACTGTG	360	
QY	361	GTTTCAGAACTTATGAGAGAAATAACCTTCTATGATTCATTTGGCGGCTGAAGTT	420	

Db 361 GTTCAGAAATCTATGAGAGAAATAACCTTCTCTATGATTGCATTGGCGGCTGGAAGTT 420
 QY 421 GGAACAGAGACAATCATCGACAATCAAAAGTCTGGAAGACTTAATTGATGCAGCTGTTT 480
 Db 421 GGAACAGAGACAATCATCGACAATCAAAAGTCTGGAAGACTTAATTGATGCAGCTGTTT 480
 QY 481 GAAGAGTCTGGGAATACAGATATAGAAGGAATCGACACAACTTAATGCATGCTATGAGGC 540
 Db 481 GAAGAGTCTGGGAATACAGATATAGAAGGAATCGACACAACTTAATGCATGCTATGAGGC 540
 QY 541 ACAGCTGCTGCTTCAATGCTGTTAACTGATGATCCAGCTCTGGGATGGGCTTCGT 600
 Db 541 ACAGCTGCTGCTTCAATGCTGTTAACTGATGATCCAGCTCTGGGATGGGCTTCGT 600
 QY 601 GGGACACATATGCAACATGCTATGATTTTAAAGCCTGATATGCTATCTGAATATCCT 660
 Db 601 GGGACACATATGCAACATGCTATGATTTTAAAGCCTGATATGCTATCTGAATATCCT 660
 QY 661 ATAGTAGATGAAAACTCTCCATAAGTGTCACTCAGTGCATTAAGACCGCTGCTACTCT 720
 Db 661 ATAGTAGATGAAAACTCTCCATAAGTGTCACTCAGTGCATTAAGACCGCTGCTACTCT 720
 QY 721 GCTCTACTGCAAAAAAGATCCATGCCCCAGTGGCAGAAAAGGGAAATGATTAAGATTTTACC 780
 Db 721 GCTCTACTGCAAAAAAGATCCATGCCCCAGTGGCAGAAAAGGGAAATGATTAAGATTTTACC 780
 QY 781 TTGAATGATTTTGGCTTCATGATCTTTTCACTACCAATATTGTAACTGCTCAAGAAATCT 840
 Db 781 TTGAATGATTTTGGCTTCATGATCTTTTCACTACCAATATTGTAACTGCTCAAGAAATCT 840
 QY 841 CTAGCTCGGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAAATGAAATGATATC 900
 Db 841 CTAGCTCGGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAAATGAAATGATATC 900
 QY 901 TATAGTGGCCTGGAAGCCTTTGGGGAATGTTAAATTAGAAGACACCTACTTTGATAGAT 960
 Db 901 TATAGTGGCCTGGAAGCCTTTGGGGAATGTTAAATTAGAAGACACCTACTTTGATAGAT 960
 QY 961 GTGGAAGAGCATTTATGAAGGCTAGCTCTGAACCTCTTCACTCAGAAAACAAAAGGCATCT 1020
 Db 961 GTGGAAGAGCATTTATGAAGGCTAGCTCTGAACCTCTTCACTCAGAAAACAAAAGGCATCT 1020
 QY 1021 TTTACTTGTATCAAAATCAAAATGAAAAATATGTACACATCTTCAGTATATGTTCCCTTGA 1080
 Db 1021 TTTACTTGTATCAAAATCAAAATGAAAAATATGTACACATCTTCAGTATATGTTCCCTTGA 1080
 QY 1081 TCTGTTCTAGACACAGTACTCACTCAGCAATTAGCAGGGAAGAGATTGGAGTGTCTTCT 1140
 Db 1081 TCTGTTCTAGACACAGTACTCACTCAGCAATTAGCAGGGAAGAGATTGGAGTGTCTTCT 1140
 QY 1141 TATGTTCTGTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAGATGCTACACCG 1200
 Db 1141 TATGTTCTGTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAGATGCTACACCG 1200
 QY 1201 GGGTCTGCTCTTGATTAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCA 1260
 Db 1201 GGGTCTGCTCTTGATTAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCA 1260
 QY 1261 AGAAGTGGTGTGGACACAGATGTCTTGGCTGAAAAACATGAAGCTCAGAGAGACACCCAT 1320
 Db 1261 AGAAGTGGTGTGGACACAGATGTCTTGGCTGAAAAACATGAAGCTCAGAGAGACACCCAT 1320
 QY 1321 CATTTGGTCAACTATATCCCGAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTATC 1380
 Db 1321 CATTTGGTCAACTATATCCCGAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTATC 1380
 QY 1381 TTAGTTAGGTTGATGAAGAACACAGAAAGACTTACGCTGGCGTCCCACTCCAAATGAT 1440
 Db 1381 TTAGTTAGGTTGATGAAGAACACAGAAAGACTTACGCTGGCGTCCCACTCCAAATGAT 1440
 QY 1441 GACACTTTGATGAAGAGAGTAGAAGCTTGTGATTCAAACATAGCAACTGAGCATATTCCA 1500
 Db 1441 GACACTTTGATGAAGAGAGTAGAAGCTTGTGATTCAAACATAGCAACTGAGCATATTCCA 1500

QY 1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCCACAGACAGAAACCTGAAGCAGCTGTC 1560
 Db 1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCCACAGACAGAAACCTGAAGCAGCTGTC 1560
 QY 1561 ATTAGTATGGGGAACATTAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGC 1620
 Db 1561 ATTAGTATGGGGAACATTAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGC 1620
 QY 1621 ATGGGTTGGGGTATGGAAAGAGTTGGAGAAATGGATATCTGGGGATTAATTAAAGGA 1680
 Db 1621 ATGGGTTGGGGTATGGAAAGAGTTGGAGAAATGGATATCTGGGGATTAATTAAAGGA 1680
 QY 1681 TTACATGTTATGTAAATTTTATATGTAGTACATGAGAGCCTGGAATGACTATCGTGTACTT 1740
 Db 1681 TTACATGTTATGTAAATTTTATATGTAGTACATGAGAGCCTGGAATGACTATCGTGTACTT 1740
 QY 1741 GGGAAAGTCTCTTTGCTCTATTTGCTGACATGCTTCTGTTGGTCTGGCCAAATGCCAA 1800
 Db 1741 GGGAAAGTCTCTTTGCTCTATTTGCTGACATGCTTCTGTTGGTCTGGCCAAATGCCAA 1800
 QY 1801 ATGTAATCTGAATGATGTTAAGGCTCTGTAAACTTCATACCTCTTTGGCCATTTGTATG 1860
 Db 1801 ATGTAATCTGAATGATGTTAAGGCTCTGTAAACTTCATACCTCTTTGGCCATTTGTATG 1860
 QY 1861 CATGATGTTGGTTTTTAAACATGATATGATGATTTGTACTTCTGTCAAGAAAGCA 1920
 Db 1861 CATGATGTTGGTTTTTAAACATGATATGATGATTTGTACTTCTGTCAAGAAAGCA 1920
 QY 1921 GAGTAATAATCTCCAAATTAATAATTTTAAACATGTAAATAAAAAAAAAAAAAAAAAA 1980
 Db 1921 GAGTAATAATCTCCAAATTAATAATTTTAAACATGTAAATAAAAAAAAAAAAAAAAAA 1980
 QY 1981 AAAAAAAAAAAAAAAAAA 2002
 Db 1981 AAAAAAAAAAAAAAAAAA 2002

RESULT 2
 US-10-193-295-1
 ; Sequence 1, Application US/10193295
 ; Patent No. 6620608
 ; GENERAL INFORMATION:
 ; APPLICANT: GONG, Fangcheng et al.
 ; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
 ; FILE REFERENCE: C1001195DIV
 ; CURRENT APPLICATION NUMBER: US/10/193, 295
 ; PRIOR FILING DATE: 2002-07-12
 ; PRIOR APPLICATION NUMBER: 08/819, 993
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 2002
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-10-193-295-1

Query Match 100.0%; Score 2002; DB 4; Length 2002;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCCTCCGACGACTCTCGGCAGTCCGGAAGTGGGTGGGCTATAAAGCTGT 60
 Db 1 CGCCTCCGACGACTCTCGGCAGTCCGGAAGTGGGTGGGCTATAAAGCTGT 60
 QY 61 AGCGAAGGAGGCGCGCGGACTGTCTTTCGTGCTCACTCCCTTCTGCTGCGG 120
 Db 61 AGCGAAGGAGGCGCGCGGACTGTCTTTCGTGCTCACTCCCTTCTGCTGCGG 120
 QY 121 CTCGTACGCTTGTCTTCAACATGCTGTGATCACTTCTTGAATGCAAGACTTGC 180

Db 121 CCGGTCACGCTTGCTCTTTCACCATGCGCTGGATCACTTCCTTGAATGCAGAAAGCTTGC 180
QY 181 TGGCCAAAAGATGTGGGAATTGTTGCCCTTGAGATCTATTTTCTTCTCAATATGTTGAT 240
Db 181 TGGCCAAAAGATGTGGGAATTGTTGCCCTTGAGATCTATTTTCTTCTCAATATGTTGAT 240
QY 241 CAAGCAGAGTTGAAAAATATGATGGTGTAGATGCTGGAAAAGTATACCATTGGGCTTGGGC 300
Db 241 CAAGCAGAGTTGAAAAATATGATGGTGTAGATGCTGGAAAAGTATACCATTGGGCTTGGGC 300
QY 301 CAGGCCAAGATGGGCTTCTGCAAGATAGAGAAGATATTAATCTCTTTGCAATGATGCTG 360
Db 301 CAGGCCAAGATGGGCTTCTGCAAGATAGAGAAGATATTAATCTCTTTGCAATGATGCTG 360
QY 361 GTTCAGAATCTTATGAGAGAAATAACCTTTCCTATGATTCATTGGGCGGCTGGAATT 420
Db 361 GTTCAGAATCTTATGAGAGAAATAACCTTTCCTATGATTCATTGGGCGGCTGGAATT 420
QY 421 GGAACAGAGACAATCATCGACAATAACAAGTCTGTGAAGACTAATTTGATGCAGCTGTTT 480
Db 421 GGAACAGAGACAATCATCGACAATAACAAGTCTGTGAAGACTAATTTGATGCAGCTGTTT 480
QY 481 GAAGAGTCTGGGAATACAGATATAGAAGGAATGCACACAATATGCATGCTATGAGGC 540
Db 481 GAAGAGTCTGGGAATACAGATATAGAAGGAATGCACACAATATGCATGCTATGAGGC 540
QY 541 ACAGCTGCTGCTTCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGATGGGCTTCT 600
Db 541 ACAGCTGCTGCTTCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGATGGGCTTCT 600
QY 601 GGGACACATATGCACATGCGCTAATGTTTTTACAAGCCTGATATGCTATCTGAATATCT 660
Db 601 GGGACACATATGCACATGCGCTAATGTTTTTACAAGCCTGATATGCTATCTGAATATCT 660
QY 661 ATAGTAGATGGAATACTCTCCATAACAGTGTACTACCTCAGTGCATTAGACCGCTGCTACT 720
Db 661 ATAGTAGATGGAATACTCTCCATAACAGTGTACTACCTCAGTGCATTAGACCGCTGCTACT 720
QY 721 GTCTACTGCAAAAAAGATCCATGCCCCAGTGGCAGAAAAGGGAAATGATTAAGATTTTACC 780
Db 721 GTCTACTGCAAAAAAGATCCATGCCCCAGTGGCAGAAAAGGGAAATGATTAAGATTTTACC 780
QY 781 TTGAATGATTTTGGCTTCATGATCTTCACTCACCATATGTAACTGGTTCAGAAATCT 840
Db 781 TTGAATGATTTTGGCTTCATGATCTTCACTCACCATATGTAACTGGTTCAGAAATCT 840
QY 841 CTAGCTCGGATGTTGCTGAATGACTTCTTAATGACCAGAAATAGATAAAATAGTATC 900
Db 841 CTAGCTCGGATGTTGCTGAATGACTTCTTAATGACCAGAAATAGATAAAATAGTATC 900
QY 901 TATAGTGGCCTGGAAGCCTTTGGGATGTTAATTAGAGAACACTACTTTGATAGAT 960
Db 901 TATAGTGGCCTGGAAGCCTTTGGGATGTTAATTAGAGAACACTACTTTGATAGAT 960
QY 961 GTGAGAGAAGGCATTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAAACAAGGCATCT 1020
Db 961 GTGAGAGAAGGCATTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAAACAAGGCATCT 1020
QY 1021 TTAATCTGTATCAAAATCAAAATGAAATATGTACACATCTTCAGTATATGTTCCCTTGA 1080
Db 1021 TTAATCTGTATCAAAATCAAAATGAAATATGTACACATCTTCAGTATATGTTCCCTTGA 1080
QY 1081 TCTGTTCTAGACAGTACTCACTCAGCAATTAGCAGGGAAGAAATTGGAGTGTCTTCT 1140
Db 1081 TCTGTTCTAGACAGTACTCACTCAGCAATTAGCAGGGAAGAAATTGGAGTGTCTTCT 1140
QY 1141 TATGTTCTGTTTGGCTGCCACTCTGTACTCTCTTAAGTCAACAAGATGCTACACCG 1200
Db 1141 TATGTTCTGTTTGGCTGCCACTCTGTACTCTCTTAAGTCAACAAGATGCTACACCG 1200
QY 1201 GGGTCTGCTCTTGATAAATAACAGCAAGTTTATGTGATCTTAATCAAGGCTTGATTCA 1260
Db 1201 GGGTCTGCTCTTGATAAATAACAGCAAGTTTATGTGATCTTAATCAAGGCTTGATTCA 1260

Db 1201 GGGTCTGCTCTTGATAAATAACAGCAAGTTTATGTGATCTTAATCAAGGCTTGATTCA 1260
QY 1261 AGAATGCTGTGGACACCAAGTGTCTTCGCTGAAAAATGAAGCTCAGAGAGACACCCAT 1320
Db 1261 AGAATGCTGTGGACACCAAGTGTCTTCGCTGAAAAATGAAGCTCAGAGAGACACCCAT 1320
QY 1321 CATTTGGTCACTATATTCGCCAGGGTTCAATAGATTCACTCTTTGAAGAAAGTGGTAC 1380
Db 1321 CATTTGGTCACTATATTCGCCAGGGTTCAATAGATTCACTCTTTGAAGAAAGTGGTAC 1380
QY 1381 TTAGTTAAGGTGATGAAAAAGCACAGAAGACTTACGCTCGGCGCTCCCAATGAT 1440
Db 1381 TTAGTTAAGGTGATGAAAAAGCACAGAAGACTTACGCTCGGCGCTCCCAATGAT 1440
QY 1441 GACACTTTGATGAAGAGTAAAGACTTGTGCATTCAACAATGCAACTGAGCATATTCCA 1500
Db 1441 GACACTTTGATGAAGAGTAAAGACTTGTGCATTCAACAATGCAACTGAGCATATTCCA 1500
QY 1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAGTGT 1560
Db 1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAGTGT 1560
QY 1561 ATTAGTATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGC 1620
Db 1561 ATTAGTATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGC 1620
QY 1621 ATGGGGTGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGGATAATTTAAAGGA 1680
Db 1621 ATGGGGTGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGGATAATTTAAAGGA 1680
QY 1681 TTACATGTTATGTAATTTTATGTGACTGACATGAGCGCTGGATGACTATCGTGTACTT 1740
Db 1681 TTACATGTTATGTAATTTTATGTGACTGACATGAGCGCTGGATGACTATCGTGTACTT 1740
QY 1741 GGGAAAGTCTCTTGTCTTAATTTGCTGACATGCTTCTGTGTGCTGGCCAATGGCAA 1800
Db 1741 GGGAAAGTCTCTTGTCTTAATTTGCTGACATGCTTCTGTGTGCTGGCCAATGGCAA 1800
QY 1801 ATGTAATCGAATGATGTTAAAGGCTCTGTAAAACTTCACTCTTTGGCCATTTGTATG 1860
Db 1801 ATGTAATCGAATGATGTTAAAGGCTCTGTAAAACTTCACTCTTTGGCCATTTGTATG 1860
QY 1861 CATGATGTTGTTTTTAAACATGCTATATGAATGTGTACTTCTGCAGAAAGAGCA 1920
Db 1861 CATGATGTTGTTTTTAAACATGCTATATGAATGTGTACTTCTGCAGAAAGAGCA 1920
QY 1921 GAGGTACTAATCTCCAATTAATAAATTTTAAACATGTAAATAAAAAAAAAAAAAA 1980
Db 1921 GAGGTACTAATCTCCAATTAATAAATTTTAAACATGTAAATAAAAAAAAAAAAAA 1980
QY 1981 AAAAAAAAAAAAAAAAAAAAAA 2002
Db 1981 AAAAAAAAAAAAAAAAAAAAAA 2002

RESULT 3
US-08-305-505-1
; Sequence 1, Application US/08305505
; Patent No. 5668001
; GENERAL INFORMATION:
; APPLICANT: Miziorko, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGUTARYL-COA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:

Fri Jun 25 07:33:50 2004

us-10-622-516-1.rni

Page 4

```

; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
FILING DATE:
; CLASSIFICATION: 435
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,040
; FILING DATE: 02 JUNE 1993
ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 65-053-9083-9
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
LENGTH: 1824 base pairs
; TYPE: nucleic acid
STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-305-505-1

```

Query Match	39.7%;	Score 794.6;	DB 1;	Length 1824;
Best Local Similarity	72.3%;	Pred. No. 4e-185;		
Matches 1139;	Conservative	0;	Mismatches 304;	Indels 132;
				Gaps 3;

Qy	144	CATGCCCTGGATCACTTCCTCTTGAATGCAGAAAGCTTGCTGGCCCAAAAGATGTGGAAATTGT	203
Db	1	CATGGCTGGGTCCTCCAGTGAACACTGAAATCCTGCTGGCCCAAGATGTGGTAATTGT	60
Qy	204	TGCCCCCTGAGATCTAATTTTCCCTCTCAATATGTGTGAATCAAGCAGAGTTGGAAAAATATGA	263
Db	61	TGCACCTGGAATCTAATTTTCCCTCTCAGTATGTGCAACCACTGAGCTGAGAGTATGA	120
Qy	264	TGGTGTAGATCGTGGAAAGTAATACCAATTGGCTTGGGCCCAGGCCAAGATGGGCTTCGCAC	323
Db	121	CGGTGTGATGCAGGCCAAATACACCAATTGGGTTAGGCCAGTCAAGATGGGCTTCGCTC	180
Qy	324	AGATAGAGAAGATATTAACTCTCTTTGGCATGACTGTGTTCAAGATCTTATGAGAGAAA	383
Db	181	TGACCGAAGAGATATCAATTCCCTCTGTTTGACTGTGTTCAAGACTTATGAGAGGAA	240
Qy	384	TAACTTTTCCCTATGATTGGCATTGGGCGGCTGGAAAGTTGGAACAGAGACAATCATGACAA	443
Db	241	CAGCTTTTCCCTATGACTGCATTGGGAGACTGGAAAGTTGGAACGAGACAATTAATTGATPA	300
Qy	444	ATCAAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATAT	503
Db	301	ATCAAAATCGGTGAAGACTGTCTGATGCAGCTAATTTGAAGAAATCTGTAATACAGATGT	360
Qy	504	AGAAGGAATGCACAACAATAATGATGCTATGAGGSCACAGCTGCTGCTTCAATGCTGT	563
Db	361	AGAAGGAATTGACAACAACCAATGCGTCTATGAGGSCACTGCTCTTTTAAATGCTAT	420
Qy	564	TAACTGGATTGAGTCCAGCTCTTTGGGATGGGC-----	595
Db	421	TAACTGGATTGAGTCCAGTCTTTGGGATGGAGCCTAATGCACCTGTGTGCTGAGACAT	480
Qy	596	-----	595
Db	481	TGCTGTGTATGCCACTGGAAATGCCAGGCCAACAGTGGAGCTGTGCTGTTGCTAATGCT	540
Qy	596	-----TTGTTGGGACACATATGACACA	617
Db	541	AGTTGGGTCAAAATGCTCTTTAATTTTGAAGAGAGATTGGGTGAACCCACATGACAGCA	600
Qy	618	TGCGTATGATTTTTCAAAGCCGTGATATGCTATCTGAATATCCTAATAGTAGAGAAAATCT	677

Db	601	TGCTTATGACTTCTATTAACCAAGATATGGTTTCTGAAATATCCTGTAGTTGATGGCAAACT	660
QY	678	CTCCATACAGTGTCTACTCAGTGCATTTAGACCGCTGCTACTCTGTCTACTGCCAAAAGAT	737
Db	661	ATCTATACAGTGTCTACTCAGTGCATTTAGACCGCTGCTATAGTGTATTATCGCAATATAAT	720
QY	738	CCATGGCCAGTGGCAGAAAGAGGGAATGATTAAGATTTTAACTTTGAATGATTTTGGCTT	797
Db	721	CCATGCCCACTGGCAAAAAGAGGGGACAGACAGAGGTTTCACTTGAATGATTTTGGATT	780
QY	798	CATGATCTTTTCACTCACCATATTTGTTAACTGGTTCAGAAATCTCTAGCTCGATGTGCT	857
Db	781	CATGATCTTTTCACTCTCCCTACTGTAACTGTATACAGAAGTCCGTGGCAAGACTGTGCT	840
QY	858	GAATGACTTCTCTTAATGACCAAGATAGAGTA--AAAATAGTATCTATAGTGGCTGGA	914
Db	841	GAATGACTTCTCTAGTGACCAAGATGACAAAACAGCAATAGTGTGTTTCAGTGTCTGGA	900
QY	915	AGCCTTTGGGGATGTTAAATTGAAGACACCTACTTTGATAGATGTGAGAGGCATT	974
Db	901	AGCTTTCAGGATGTAAGCTTGAAGATACATATTTGATAGGATGTGAAAAAGCTTT	960
QY	975	TATGAAGCCTAGCTCTGAACCTTTCAGTCAGAAAACAAAGGCATCTTACTGTATCAAA	1034
Db	961	TATGAAGCTAGTGACAGAGCTCTTCAATCAGAAAACCAAAGCTTCTTACTGTATCAAA	1020
QY	1035	TCAAAATGGAATATGTACACATCTTCAGTATATGTGTTCCCTTGCATCTGTTTAGACA	1094
Db	1021	TCAGAAATGGAACATGTACACGCTTCAGTCTACGGTTGCCCTTGCTTCTTCTTAGCCCA	1080
QY	1095	GTACTCACTCAGCAATTAGCAGGGAGAGAAATTGGAGTGTGTTTCTTATGTTCTGTGTTT	1154
Db	1081	GTACTCTTCAGAGCACTTCGACAGACAAAGAAATCAGTGAGTTCTCATATGGCTCTGTGTTT	1140
QY	1155	GGCTGCCACTCTGTACTCTCTTAAAGTCACACAAAGATGCTACACCGGGCTGCTCTTGA	1214
Db	1141	TGCTGCTACGCTGTATTCCATCAGAGTTACACAGSATGCCACTCGGTCTCGCGTTGA	1200
QY	1215	TAAATATACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTTCAAGAACTGTTGCGC	1274
Db	1201	CAAAATATCTGCTAGCCTTCTGATCTTAAAGCAAGACTTGACTCAGAAATGCAATTGC	1260
QY	1275	ACCAGATGCTTTCGCTGTAACAATGAAGCTCAGAGAGGACACCCATCATTTTGGTCAACTA	1334
Db	1261	ACCTGATGCTTTTGTCTGAATAACATGAAGATTAGACAGAGACACATCACTTGGCCACTA	1320
QY	1335	TATTCGCCAGGGTTCAATAGATTCACCTTTTGAAGGAACGTGTACTTATGTTAGGTTGA	1394
Db	1321	TATTCACAGTGTTCAGTAGAAGATCTCTTTGAGGGAACATGTATCTTGTGCGTGTGA	1380
QY	1395	TGAAAAAGCACAGAAGACTTACGCTCGCGTCCCACTCCAATGATGACACTTTGGATGA	1454
Db	1381	TGAAAAACACAGAGAACATATGACAGACGCCCACTTATGGGTGATGACCCCTGAGGC	1440
QY	1455	AGGAGTAGGACTTGTGCAATTCAAAATAGCAATGAGCATATTTCCAAAGCCCTGCCAAGAA	1514
Db	1441	AGGAGTTGAAGTTGTCCACCACGAGCATTTGTTCAATGAGCACATCCCAAGCCCTGCTAAGAA	1500
QY	1515	AGTACCAAGACTCCCTGCCACAGCAG--CAGAACTGAAGCAGCTGTTCATTAGTAATGG	1571
Db	1501	AGTGCCAAGATCCCTGCACAAACAAGAACTTGAAAGCGCTTACTGTTGCCATTTCCAAATGG	1560
QY	1572	GGAACATTAAAGATAC	1586
Db	1561	GGTGCAATTAAAGATAC	1575

RESULT 4
US-09-819-993-3
; Sequence 3, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.


```
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 28001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(28001)
; OTHER INFORMATION: n = A,T,C or G
US-09-819-993-3
```

```
Query Match      23.8%; Score 476; DB 4; Length 28001;
Best Local Similarity 95.2%; Pred. No. 1.4e-106;
Matches 491; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
```

```
QY 1479 CATAGCAACTGAGCATATTCCAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGC 1538
    |||||
Db 24597 CCTTCATTTTAGCATATATCCAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGC 24656

QY 1539 AGCAGAACTGAAGCAGCTGTCAATTAGTAATGGGGAACATTAAAGATACTCTGTGAGGTGC 1598
    |||||
Db 24657 AGCAGAACTGAAGCAGCAGCTGTCAATTAGTAATGGGGAACATTAAAGATACTCTGTGAGGTGC 24716

QY 1599 AAGACTTCAGGGTGGGGTGGGCATGGGGTGGGGTATGGGAACAGTTGGAGGAATGGGAT 1658
    |||||
Db 24717 AAGACTTCAGGGTGGGGTGGGCATGGGGTGGGGTATGGGAACAGTTGGAGGAATGGGAT 24776

QY 1659 ATCTGGGGATAATTTTAAAGATTACATGTTATGTAATTTTATGTGACTGACATGGAG 1718
    |||||
Db 24777 ATCTGGGGATAATTTTAAAGATTACATGTTATGTAATTTTATGTGACTGACATGGAG 24836

QY 1719 CCTGGATGACTATCGTGTACTTGGGAAAGTCTCTTGTGCTTAATTGCTGACATGCTTCT 1778
    |||||
Db 24837 CCTGGATGACTATCGTGTACTTGGGAAAGTCTCTTGTGCTTAATTGCTGACATGCTTCT 24896

QY 1779 GTTGTGCTCTGGCCATGCGCAATGTACTCGAATGATGTTAAGGGCTCTGTAAACCTTCA 1838
    |||||
Db 24897 GTTGTGCTCTGGCCATGCGCAATGTACTCGAATGATGTTAAGGGCTCTGTAAACCTTCA 24956

QY 1839 TACCTCTTTGGCCATTGTATGCATGATGTTGGTTTAAACATGSTATATGAATTGT 1898
    |||||
Db 24957 TACCTCTTTGGCCATTGTATGCATGATGTTGGTTTAAACATGSTATATGAATTGT 25016

QY 1899 GTAAGTCTGTGAGAAAGCAGAGTACTAATCTCCAATTAAAAAAATTTTAAACATGT 1958
    |||||
Db 25017 GTAAGTCTGTGAGAAAGCAGAGTACTAATCTCCAATTAAAAAAATTTTAAACATGT 25076

QY 1959 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1994
    |||||
Db 25077 AAGAAATTTTGTACTTTGAACAACAAGATTACAGAAA 25112
```

```
RESULT 5
US-10-193-295-3
; Sequence 3, Application US/101933295
; Patent No. 6620608
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195DIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
```

```
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 28001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(28001)
; OTHER INFORMATION: n = A,T,C or G
US-10-193-295-3
```

```
Query Match      23.8%; Score 476; DB 4; Length 28001;
Best Local Similarity 95.2%; Pred. No. 1.4e-106;
Matches 491; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
```

```
QY 1479 CATAGCAACTGAGCATATTCCAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGC 1538
    |||||
Db 24597 CCTTCATTTTAGCATATATCCAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGC 24656

QY 1539 AGCAGAACTGAAGCAGCTGTCAATTAGTAATGGGGAACATTAAAGATACTCTGTGAGGTGC 1598
    |||||
Db 24657 AGCAGAACTGAAGCAGCAGCTGTCAATTAGTAATGGGGAACATTAAAGATACTCTGTGAGGTGC 24716

QY 1599 AAGACTTCAGGGTGGGGTGGGCATGGGGTGGGGTATGGGAACAGTTGGAGGAATGGGAT 1658
    |||||
Db 24717 AAGACTTCAGGGTGGGGTGGGCATGGGGTGGGGTATGGGAACAGTTGGAGGAATGGGAT 24776

QY 1659 ATCTGGGGATAATTTTAAAGATTACATGTTATGTAATTTTATGTGACTGACATGGAG 1718
    |||||
Db 24777 ATCTGGGGATAATTTTAAAGATTACATGTTATGTAATTTTATGTGACTGACATGGAG 24836

QY 1719 CCTGGATGACTATCGTGTACTTGGGAAAGTCTCTTGTGCTTAATTGCTGACATGCTTCT 1778
    |||||
Db 24837 CCTGGATGACTATCGTGTACTTGGGAAAGTCTCTTGTGCTTAATTGCTGACATGCTTCT 24896

QY 1779 GTTGTGCTCTGGCCATGCGCAATGTACTCGAATGATGTTAAGGGCTCTGTAAACCTTCA 1838
    |||||
Db 24897 GTTGTGCTCTGGCCATGCGCAATGTACTCGAATGATGTTAAGGGCTCTGTAAACCTTCA 24956

QY 1839 TACCTCTTTGGCCATTGTATGCATGATGTTGGTTTAAACATGSTATATGAATTGT 1898
    |||||
Db 24957 TACCTCTTTGGCCATTGTATGCATGATGTTGGTTTAAACATGSTATATGAATTGT 25016

QY 1899 GTAAGTCTGTGAGAAAGCAGAGTACTAATCTCCAATTAAAAAAATTTTAAACATGT 1958
    |||||
Db 25017 GTAAGTCTGTGAGAAAGCAGAGTACTAATCTCCAATTAAAAAAATTTTAAACATGT 25076

QY 1959 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1994
    |||||
Db 25077 AAGAAATTTTGTACTTTGAACAACAAGATTACAGAAA 25112
```

```
RESULT 6
US-09-370-838-245
; Sequence 245, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 245
; LENGTH: 615
```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105)
; OTHER INFORMATION: n=A,T,C or G
US-09-370-838-245

Query Match 21.8%; Score 436.4; DB 4; Length 615;
Best Local Similarity 99.1%; Pred. No. 1.4e-97;
Matches 459; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 134 GCTCTTTCACCATGCGCTGGATCAGTCTTCTTGAATGCAGAAAGCTTGCGCCAAAAGATG 193
Db 92 GCTCTTTCACCAT-NCTGGATCAGTCTTCTTGAATGCAGAAAGCTTGCGCCAAAAGATG 150
QY 194 TGGGAATTGTTGCCCTTGAGATCTATTTCTCTTCAATATGTTGATCAAGCAGAGTTGG 253
Db 151 TGGGAATTGTTGCCCTTGAGATCTATTTCTCTTCAATATGTTGATCAAGCAGAGTTGG 210
QY 254 AAAAATATGATGTTGATGCTGGAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG 313
Db 211 AAAAATATGATGTTGATGCTGGAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG 270
QY 314 GCTTCTGCACAGATAGAGAATATTACTCTCTTGATGACGTGGTTCAGAACTTTA 373
Db 271 GCTTCTGCACAGATAGAGAATATTACTCTCTTGATGACGTGGTTCAGAACTTTA 330
QY 374 TGGAGAAATAAATCACTTCTCTATGATTCATTTGGCGGGCTGGAAGTTGGAACAGACAA 433
Db 331 TGGAGAAATAAATCACTTCTCTATGATTCATTTGGCGGGCTGGAAGTTGGAACAGACAA 390
QY 434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTGAAGAGTCTGGGA 493
Db 391 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTGAAGAGTCTGGGA 450
QY 494 ATACAGATATAGAAAGGAATGCACCAACTAATGATGCTATGAGGACAGAGCTGCTGTCT 553
Db 451 ATACAGATATAGAAAGGAATGCACCAACTAATGATGCTATGAGGACAGAGCTGCTGTCT 510
QY 554 TCAATGC-TGTTAACTGGATTGATGCCAGCTCTTGGGATGGC 595
Db 511 TCAATGCTTGTTAACTGGATTGATGCCAGCTCTTGGGATGGAC 553

RESULT 7
US-09-370-838-128
; Sequence 128, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 128
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-370-838-128

Query Match 20.5%; Score 410; DB 4; Length 500;
Best Local Similarity 100.0%; Pred. No. 3.9e-91;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 GCTCTTTCACCATGCGCTGGATCAGTCTTCTTGAATGCAGAAAGCTTGCGCCAAAAGATG 193
Db 91 GCTCTTTCACCATGCGCTGGATCAGTCTTCTTGAATGCAGAAAGCTTGCGCCAAAAGATG 150
QY 194 TGGGAATTGTTGCCCTTGAGATCTATTTCTCTTCAATATGTTGATCAAGCAGAGTTGG 253
Db 151 TGGGAATTGTTGCCCTTGAGATCTATTTCTCTTCAATATGTTGATCAAGCAGAGTTGG 210
QY 254 AAAAATATGATGTTGATGCTGGAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG 313
Db 211 AAAAATATGATGTTGATGCTGGAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG 270
QY 314 GCTTCTGCACAGATAGAGAATATTACTCTCTTTCATGACTGTGTTCAGAACTTTA 373
Db 271 GCTTCTGCACAGATAGAGAATATTACTCTCTTTCATGACTGTGTTCAGAACTTTA 330
QY 374 TGGAGAAATAAATCACTTCTCTATGATTCATTTGGCGGGCTGGAAGTTGGAACAGACAA 433
Db 331 TGGAGAAATAAATCACTTCTCTATGATTCATTTGGCGGGCTGGAAGTTGGAACAGACAA 390
QY 434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTGAAGAGTCTGGGA 493
Db 391 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTGAAGAGTCTGGGA 450
QY 494 ATACAGATATAGAAAGGAATGCACCAACTAATGATGCTATGAGGACACA 543
Db 451 ATACAGATATAGAAAGGAATGCACCAACTAATGATGCTATGAGGACACA 500

RESULT 8
US-09-401-064-187
; Sequence 187, Application US/09401064
; Patent No. 6623923
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeline Joy
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C2
; CURRENT APPLICATION NUMBER: US/09/401,064
; CURRENT FILING DATE: 1999-09-22
; SOFTWARE: FastSeq for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 371
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 187
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-401-064-187

Query Match 20.0%; Score 401; DB 4; Length 506;
Best Local Similarity 100.0%; Pred. No. 6.2e-89;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 134 GCTCTTTCACCATGCGCTGGATCAGTCTTCTTGAATGCAGAAAGCTTGCGCCAAAAGATG 193
Db 106 GCTCTTTCACCATGCGCTGGATCAGTCTTCTTGAATGCAGAAAGCTTGCGCCAAAAGATG 165
QY 194 TGGGAATTGTTGCCCTTGAGATCTATTTCTCTTCAATATGTTGATCAAGCAGAGTTGG 253
Db 166 TGGGAATTGTTGCCCTTGAGATCTATTTCTCTTCAATATGTTGATCAAGCAGAGTTGG 225
QY 254 AAAAATATGATGTTGATGCTGGAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG 313
Db 226 AAAAATATGATGTTGATGCTGGAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGG 285
QY 314 GCTTCTGCACAGATAGAGAATATTACTCTCTTTCATGACTGTGTTCAAGAACTTTA 373
Db 286 GCTTCTGCACAGATAGAGAATATTACTCTCTTTCATGACTGTGTTCAAGAACTTTA 345

OY	374	TGAGAGAATAACCTTTCCATGATTGCATTGGGGCGGCTGGAAGTTGGAACAGACA	433
Db	346	TGAGAGAAATAACCTTTCCTATGATTGCATTGGCGGCTGGAAGTTGGAACAGACA	405
OY	434	TCATGCACCAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGA	493
Db	406	TCATGCACCAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGA	465
OY	494	ATACAGATATAGAAGGATCGACACACTAATGCATGCTAT	534
Db	466	ATACAGATATAGAAGGATCGACACACTAATGCATGCTAT	506

```

RESULT 9
US-09-833-381-1645
; Sequence 1645, Application US/098333381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1645
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(448)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1645

```

Query Match	11.2%;	Score 224;	DB 4;	Length 448;
Best Local Similarity	82.1%;	Pred. No. 1.4e-45;		
Matches 262; Conservative	0;	Mismatches 54;	Indels 3;	Gaps 2;

QY	1142	ATGTTCTGTGGTTGGCTGCCACTCTGTACTCTTTAAAGTCACACAAGATGCTACACCGG	1201
Dp	8	ATGGTCTGGINTGGCTGCCACTCTNTANTCTTTANAATTCACACNNAATGCTNCACCGG	67
QY	1202	GGTCTGCTCTTGATATAAATAACAGCAAGTTTANGTGATCTTAAATCAAGGCTTGATTCAA	1261
Dp	68	GGTCTGCTCTTGATACAATNNATCAANTTNATGTGATCTTAAATCAANGCTTGATTGCA	127
QY	1262	GAACTGGTGTGGACACAGATGTCTTGCTGAAAAACATGAAGCTCAGAGAGACACCCATC	1321
Dp	128	TAACTGGTGTGGCAACATATGTCTTCNCNGANAANNNTGAANCTCATATANGACACCCATC	187
QY	1322	ATTTGGTCACTATATTTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGTACT	1381
Dp	188	ATTTGGTCACTATATTTCCCCATGGTTCAATACATTCCTTTGAAGAACNTGGTNTCT	247
QY	1382	TAGTTA-GGGTGGATGAATAAGCACAGAGAATTAACGCTTGGCGCTCCC--ACTCCAAATG	1438
Dp	248	TACTTNTGGTTGGATGATGAATGAACNTANAAAGAAATTNCGCTGGAGTCCCCCACTCCCNAAATN	307
QY	1439	ATGACACTTTGGATGAAGG	1457
Dp	308	ATGACACTTTGGATCAAG	326

RESULT 10
US-09-833-381-1648
; Sequence 1648, Application US/098333381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.

```

; TITLE OF INVENTION: No. 6672186e1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1648
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(472)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1648

```

Query Match	10.5%;	Score 210.4;	DB 4;	Length 472;
Best Local Similarity	70.1%;	Pred. No. 3e-42;		
Matches 333; Conservative	0;	Mismatches 129;	Indels 13;	Gaps 4;

QY	1270	GTGGCACCAGATGTCCTTCGCTGAAAACATGAA-GCTCAGAGAGACACCCATCATTTGGT	1328
Db	8	GTGGCACCAGATGTCCTTCGCTGAAAACATGATGCTCAGAGAGACACACATCATTAGC	67
QY	1329	CAACTATATATCCCAAGGGTTCATATAGATTCACTCTTTGAAGGAACGTGTACTTAACTTAA	1388
Db	68	CAACTATATATCCCA-TGTTCAATAMACTCACTCTTTGAAGGAACGTGTATCTGGTCAG	126
QY	1389	GGTGATGAAAAGCACAAGAAGACTTACGCTCGGCGTCCCACTCCAATGATGACACTTT	1448
Db	127	AGTGATGAAAACACAGAAGACTTACGCGCGGCCCTTCACAATGACCACTGTTT	186
QY	1449	GGATGAAGAGTAGGACTTGTGCATTCAACACATAGCAACTGAGCATATTTCCAAGCCCTGC	1508
Db	187	GGATGAAGAGATGNNCTCGTGATAGTAACAAGCAACAGAGCATATTTCCAAGCCCTGC	246
QY	1509	CAAGAAAGTACCAAGACTCCCTGCCACAGACAGCAAACTGAAGCAGCTGTCAATAGTAA	1568
Db	247	TAAGAAAGTGCCAGACTCCCTGCACACCTCGGCCGAACTTGAATCACNCTGTCAATCAAT	306
QY	1569	TGGGGAACATTAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGCATGGGGTG	1628
Db	307	AACGGGAGCACTGAGAGACTGTGTGGCCTTTACAG-----AGGCTCGGNGCCGGAAT	357
QY	1629	GGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGGATAATTTAAAGATTTACATGT	1688
Db	358	GGGGTATGGAAACCGGCTTGGAGGAATGATGTCTTTGGACAATTTTGCAGATTTATNGT	417
QY	1689	T-ATGTAATTTTATGTGACTGACATGAGCCTTGGATGACTATCGTACTTG 1741	
Db	418	TTGCTTAAAAATGNNATGGAACCTGCACAGANCCAGAAAAGCTNTTGNGTTTTTG 472	

RESULT 11
 US-08-998-416-730
 ; Sequence 730, Application US/08998416
 ; Patent No. 6239264
 ; GENERAL INFORMATION:
 ; APPLICANT: Philippsen, Peter
 ; APPLICANT: Pohlmann, Rainer
 ; APPLICANT: Steiner, Sabine
 ; APPLICANT: Mohr, Christine
 ; APPLICANT: Wendland, Jurgen
 ; APPLICANT: Knechtle, Philipp
 ; APPLICANT: Rebischung, Corinne
 ; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 1152
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: No. 6239264artis Corporation
 ; STREET: 3054 Cornwallis Road


```

; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGCI976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 730:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1476UP
;
US-08-998-416-730
```

```

Query Match      6.8%; Score 135.8; DB 3; Length 635;
Best Local Similarity 62.6%; Pred. No. 6.4e-24;
Matches 231; Conservative 0; Mismatches 132; Indels 6; Gaps 1;

QY      229 CAATATGTTGATCAAGCAGAGTTGAAAAATATGATGTTAGTCTGTAAGTATACC 288
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      20  CAGTGCCTGAACCAAGAGGCATTGAGGCGGTATGACGGCGTGTGCAAGGCAAGTACACT 79
QY      289 ATTGCTTGGGCCAGGCCAAGATGGGCTTTCGACAGATAGAGAATATTAATCTCTTT 348
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      80  ATCGGCTTGGGCCAGACCAACATGAGCTTTGTGAACGACCGCGAAGACATCTCGATG 139
QY      349 TGCATGACTGTGGTTCAGAACTTATGAGAGAAATTAACCTTCTATGATTCATTGGG 408
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      140 TGTTTGACCGCGTGTCTGAACTTGATGAAGACTACGATATCAAGCCGGAAGCATCGGC 199
QY      409 CGGCTGGAAGTTGGAACAGAGCAATCATCGACAAATCAAACTCTGTGAAGACTAATTGG 468
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      200 CGCCTCGAGGTGGGTACCGGAGACGTTGCTTGACAAAGTCCGGAAGTCTATTTTG 259
QY      469 ATGCAGCTGTTGAAGAGTCTGGGAATACAGATATAGAGAATCGACACAATAATGCA 528
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      260 ATGCAGTTGTTCGGCGA-----GAACACCGAAGCTTGAGGGTGTGATACCGTGAACGCC 313
QY      529 TGCTATGAGGACACAGCTGTCTTCAATGCTGTTAAGTGAATGAGTCCAGCTCTTGG 588
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      314 TGCTATGGCGGTACTAACCGGTGTTTAAGTCTTGAAGTGAATGAGTCCAGTTCGTGG 373
QY      589 GATGGGCTT 597
      ||| ||| |||
Db      374 GACGGTCTGT 382
```

```

RESULT 12
US-09-833-381-744
; Sequence 744, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
```

```

; TITLE OF INVENTION: No. 6672186e1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 744
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(307)
; OTHER INFORMATION: n = A,T,C or G
;
US-09-833-381-744
```

```

Query Match      5.5%; Score 110; DB 4; Length 307;
Best Local Similarity 56.9%; Pred. No. 9.9e-18;
Matches 173; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY      775 TTTAACCTTGAATGATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAGTGTTCAG 834
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2  TNCACCCCTNGACGATTTACAGTACATGANTCTTTCATACANCTTTGCAAGNNGTCCAG 61
QY      835 AAATCTTACGCTCGGATGTTGCTCTGATGACTTCTTAATGACCAAGATAGATAAAT 894
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      62  ANGNCTCTNGCTCGNCTGATNTTCAATGACTTNGTGTGAGNAGCAGTGACACAAACC 121
QY      895 AGTATCTATAGTGGCCTGGAAGCCTTTGGGAGTGTAAATTAGAAGACACCTACTTGTAT 954
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      122 AGCTTATATAGGGGGCTNGAGGCTTTCGGGGGGCTAAAGCTGGANGACACCTNCAACCAAC 181
QY      955 AGAGATGTGAGAGAGGCATTTATGAAGGCTAGCTCTGAACCTCTTCAGTCAGAAACAAAG 1014
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      182 AAGGACCTGTGNTAAAGCACTTCTANAGGCTTNTNAGACATGTTGCAAGAAANACCAAG 241
QY      1015 GCATCTTTACTTGTATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCC 1074
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      242 GCTTCCCTTTACCTCTCCACTCAACAATNGAACATGTNCACTNACCCTGTACGGGTNC 301
QY      1075 CTTG 1078
      |||
Db      302 CTGG 305
```

```

RESULT 13
US-09-306-595C-1
; Sequence 1, Application US/09306595C
; Patent No. 6284506
; GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: OJIMA, Kazuyuki
; TITLE OF INVENTION: ISOPRENOID PRODUCTION
; FILE REFERENCE: ISOPRENOID PRODUCTION
; CURRENT APPLICATION NUMBER: US/09/306,595C
; CURRENT FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 98108210
; PRIOR FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4775
; TYPE: DNA
; ORGANISM: Phaffia rhodozyma
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1239)..(1240)
; OTHER INFORMATION: EXPERIMENTAL
; NAME/KEY: exon
; LOCATION: (1305)..(1361)
```

```

; NAME/KEY: intron
; LOCATION: (1362)..(1504)
; NAME/KEY: exon
; LOCATION: (1505)..(1522)
; NAME/KEY: intron
; LOCATION: (1523)..(1699)
; NAME/KEY: exon
; LOCATION: (1700)..(1826)
; NAME/KEY: intron
; LOCATION: (1827)..(1920)
; NAME/KEY: exon
; LOCATION: (1921)..(2277)
; NAME/KEY: intron
; LOCATION: (2278)..(2351)
; NAME/KEY: exon
; LOCATION: (2352)..(2409)
; NAME/KEY: intron
; LOCATION: (2410)..(2497)
; NAME/KEY: exon
; LOCATION: (2498)..(2504)
; NAME/KEY: intron
; LOCATION: (2505)..(2586)
; NAME/KEY: exon
; LOCATION: (2587)..(2768)
; NAME/KEY: intron
; LOCATION: (2769)..(2851)
; NAME/KEY: exon
; LOCATION: (2852)..(2891)
; NAME/KEY: intron
; LOCATION: (2892)..(2985)
; NAME/KEY: exon
; LOCATION: (2986)..(3240)
; NAME/KEY: intron
; LOCATION: (3241)..(3325)
; NAME/KEY: exon
; LOCATION: (3326)..(3493)
; NAME/KEY: intron
; LOCATION: (3494)..(3601)
; NAME/KEY: exon
; LOCATION: (3602)..(3768)
; NAME/KEY: polyA_site
; LOCATION: (4043)..(4044)
US-09-306-595C-1
```

```

Query Match      5.2%; Score 103.6; DB 3; Length 4775;
Best Local Similarity 63.0%; Pred. No. 1.1e-15;
Matches 160; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY      340 AACTCTTTGCATGACTGTGCTCAGAACTTATGAGAGAAATAACCTTTCCTATGAT 399
      |||||
Db      1902 AATTCATTGATTGTTAGCTGTTCCGGTCTTCTATCAAAAGTACAACGTTGATCCCAAG 1961

QY      400 TGCATTGGGCGGCTGGAAGTTGAAACAGACAAATCATGCACAAATCAAAGTCTGTGAG 459
      |||||
Db      1962 TCAATCGGTGGAATTGATGTCGAACTGAGTCCATCATGCAAAAGTCCAAATCTGTCAAG 2021

QY      460 ACTAATTGATGACGCTGTTGAAGAGTCTGGGAATACAGATATAGAAGGAATGCACACA 519
      |||||
Db      2022 ACAGTCCTTATGACTGTTCGAGTCCACGGCAACACAGATATTGAGGGTATCGACTCC 2081

QY      520 ACTAATGCATGCTATGAGGACACAGCTGCTGTCTTCAATGCTGTTAACTGATTGAGTCC 579
      |||||
Db      2082 AAGAATGCCCTGCTACGGGTTCTAACCGCGGCCCTGTCAATGCCGTCAACTGGATCGAGTCA 2141

QY      580 AGCTCTTGGGATGG 593
      |||||
Db      2142 TCCTCTTGGGACGG 2155
```

```

RESULT 14
US-09-925-388-1
; Sequence 1, Application US/09925388
; Patent No. 6586202
```

```

; GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: OJIMA, Kazuyuki
; APPLICANT: SETOGUCHI, Yutaka
; TITLE OF INVENTION: ISOPRENOID PRODUCTION
; FILE REFERENCE: ISOPRENOID PRODUCTION
; CURRENT APPLICATION NUMBER: US/09/925,388
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/306,595
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4775
; TYPE: DNA
; ORGANISM: Phaffia rhodozyma
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1239)..(1240)
; OTHER INFORMATION: EXPERIMENTAL
; NAME/KEY: exon
; LOCATION: (1305)..(1361)
; NAME/KEY: intron
; LOCATION: (1362)..(1504)
; NAME/KEY: exon
; LOCATION: (1505)..(1522)
; NAME/KEY: intron
; LOCATION: (1523)..(1699)
; NAME/KEY: exon
; LOCATION: (1700)..(1826)
; NAME/KEY: intron
; LOCATION: (1827)..(1920)
; NAME/KEY: exon
; LOCATION: (1921)..(2277)
; NAME/KEY: intron
; LOCATION: (2278)..(2351)
; NAME/KEY: exon
; LOCATION: (2352)..(2409)
; NAME/KEY: intron
; LOCATION: (2410)..(2497)
; NAME/KEY: exon
; LOCATION: (2498)..(2504)
; NAME/KEY: intron
; LOCATION: (2505)..(2586)
; NAME/KEY: exon
; LOCATION: (2587)..(2768)
; NAME/KEY: intron
; LOCATION: (2769)..(2851)
; NAME/KEY: exon
; LOCATION: (2852)..(2891)
; NAME/KEY: intron
; LOCATION: (2892)..(2985)
; NAME/KEY: exon
; LOCATION: (2986)..(3240)
; NAME/KEY: intron
; LOCATION: (3241)..(3325)
; NAME/KEY: exon
; LOCATION: (3326)..(3493)
; NAME/KEY: intron
; LOCATION: (3494)..(3601)
; NAME/KEY: exon
; LOCATION: (3602)..(3768)
; NAME/KEY: polyA_site
; LOCATION: (4043)..(4044)
US-09-925-388-1
```

```

Query Match      5.2%; Score 103.6; DB 4; Length 4775;
Best Local Similarity 63.0%; Pred. No. 1.1e-15;
Matches 160; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY      340 AACTCTTTGCATGACTGTGCTCAGAACTTATGAGAGAAATAACCTTTCCTATGAT 399
      |||||
Db      1902 AATTCATTGATTGTTAGCTGTTCCGGTCTTCTATCAAAAGTACAACGTTGATCCCAAG 1961
```

QY	400	TGCATTGGGCGGCTGGAAGTTGGAAACAGACAGACAATCATCGACCAATCAAAGTCTGTGAAG	459
Db	1962	TCAATCGGTGGAATTGATGTCCGGAAGTAGTCATCATTGACAAGTCCAATCTGTCAAG	2021
QY	460	ACTAATTGATGCAGCTGTTTGAAGAAGTCTGGGAATACAGATATAGAAGAATCGACACA	519
Db	2022	ACAGTCCCTATGACTTGTTCGAGTCCCACGGCAACA CAGATATTTAGGGTATCGACTCC	2081
QY	520	ACTAATGCATGCTATGAGGACACAGCTGCTGTTTTCATATGCTGTAACTGATTGAGTCC	579
Db	2082	AAGAATGCCCTGCTACGGTTCATCCGCGCCCTGTTCAATGCCCCAATGAGATCGAGTCA	2141
QY	580	AGCTCTTGGGATGG	593
Db	2142	TTCTTCTTGGGACGG	2155

RESULT 15

```

US-09-313-294A-4313
; Sequence 4313, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313, 294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 4313
; LENGTH: 305
;
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700348214H1
US-09-313-294A-4313

```

Query Match	4.3%;	Score 86.6;	DB 4;	Length 305;
Best Local Similarity	66.1%;	Pred. No. 5.3e-12;		
Matches 125; Conservative	0;	Mismatches 64;	Indels 0;	Gaps 0;

OY		414	GGAAGTTGGAACACAGACAATCATCGAACAACTCTGTGAAGACTAATTGGATGCA	473
Dd		1	GGAGGTTGGAAAGTGAACACAGTTATAGACAAAAGTAATCCATCAAACCTTGCGCTGATGCA	60
OY		474	GCTGTTTGAAGAGTCCTGGGAATACAGATATAGAAGAAATCGACACAATAATGCATGCTA	533
Dd		61	GATCTTTGAGGAAGTGGTAAACAGTAGACATTTGAAGGTGTTGACTCCAGCAATGCATGTTA	120
OY		534	TGGAGGCACAGCTGCTGTCTTCCAATGCTGTTAACTGSAATTGAGTCCAGCTCTTGGGATGG	593
Dd		121	TGGCGGGACAGACCCTACTGAAATGTGTGAATTGGGTGAAAAGTAATTCCTGGGATGG	180
OY		594	GCTTCGTGG	602
Dd		181	CCGTTATGG	189

Search completed: June 24, 2004, 11:17:11
Job time : 175 secs

QY 181 TGGCCAAAAGATGTGGGAATTGTTGCCCTTGAGATCTATTTTCTTCTCAATATGTTGAT 240
Db 181 TGGCCAAAAGATGTGGGAATTGTTGCCCTTGAGATCTATTTTCTTCTCAATATGTTGAT 240
QY 241 CAAGCAGAGTTGAAAAAATATGATGGTGTAGATGCTGGAAGATATACCATGGCTTGGGC 300
Db 241 CAAGCAGAGTTGAAAAAATATGATGGTGTAGATGCTGGAAGATATACCATGGCTTGGGC 300
QY 301 CAGGCCAAGATGGGCTTCTGCACAGATGAGAAGATATTAATCTCTTTGCATGACTGTG 360
Db 301 CAGGCCAAGATGGGCTTCTGCACAGATGAGAAGATATTAATCTCTTTGCATGACTGTG 360
QY 361 GTTCAGAACTTTATGGAGAGAAATAACTTTCTCTATGATTCATTTGGCGGCTGGAAGTT 420
Db 361 GTTCAGAACTTTATGGAGAGAAATAACTTTCTCTATGATTCATTTGGCGGCTGGAAGTT 420
QY 421 GGAACAGAGACAATCATCGACAATCAAAGTCTGTGAGAATAATTTGATGCAGCTGTTT 480
Db 421 GGAACAGAGACAATCATCGACAATCAAAGTCTGTGAGAATAATTTGATGCAGCTGTTT 480
QY 481 GAAGAGTCTGGGAATACGATATAGAGAGAAATCGACACAATAATGCATGCTATGAGAGC 540
Db 481 GAAGAGTCTGGGAATACGATATAGAGAGAAATCGACACAATAATGCATGCTATGAGAGC 540
QY 541 ACAGCTGCTGCTTCAATGCTGTTAATGGAATTGAGTCCAGCTCTTGGAGTGGCTTCGT 600
Db 541 ACAGCTGCTGCTTCAATGCTGTTAATGGAATTGAGTCCAGCTCTTGGAGTGGCTTCGT 600
QY 601 GGGACACATATGCAACATGCTATGATTTTACAAGCTGATATGCTATGTAATATCCT 660
Db 601 GGGACACATATGCAACATGCTATGATTTTACAAGCTGATATGCTATGTAATATCCT 660
QY 661 ATAGTAGATGAAAACTCTCCATACAGTGCCTCACTCAGTGCATTAGACCGCTGCTACTCT 720
Db 661 ATAGTAGATGAAAACTCTCCATACAGTGCCTCACTCAGTGCATTAGACCGCTGCTACTCT 720
QY 721 GTCTACTGCAAAAAAGATCATGCCCCAGTGGCAGAAAAGGGAATGATAAAGATTTTACC 780
Db 721 GTCTACTGCAAAAAAGATCATGCCCCAGTGGCAGAAAAGGGAATGATAAAGATTTTACC 780
QY 781 TTGAATGATTTTGGCTTCATGATCTTTCACTCAACCATATTTGTAATCGGTTCAGAATCT 840
Db 781 TTGAATGATTTTGGCTTCATGATCTTTCACTCAACCATATTTGTAATCGGTTCAGAATCT 840
QY 841 CTAGCTCGGATGTTGCTGTAATGACTTCCCTTAATGACCAAGAAATAGAGATAAAATAGTATC 900
Db 841 CTAGCTCGGATGTTGCTGTAATGACTTCCCTTAATGACCAAGAAATAGAGATAAAATAGTATC 900
QY 901 TATAGTGGCTGGAGCCCTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGAT 960
Db 901 TATAGTGGCTGGAGCCCTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGAT 960
QY 961 GTGAGAGGCAATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAGGCATCT 1020
Db 961 GTGAGAGGCAATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAGGCATCT 1020
QY 1021 TTACTTGTATCAAATCAAAAATGGAATATATGTACACATCTTCAGTATATGTTCCCTTGCA 1080
Db 1021 TTACTTGTATCAAATCAAAAATGGAATATATGTACACATCTTCAGTATATGTTCCCTTGCA 1080
QY 1081 TCTGTTCTAGCAGACTACTCACTCAGCAATTTAGCAGGGGAAGAATTGGAGTGTCTTCT 1140
Db 1081 TCTGTTCTAGCAGACTACTCACTCAGCAATTTAGCAGGGGAAGAATTGGAGTGTCTTCT 1140
QY 1141 TATGTTCTGGTTTGGCTGCCACTCTGTACTCTTTAAAGTCAACAAGATGTTACACCG 1200
Db 1141 TATGTTCTGGTTTGGCTGCCACTCTGTACTCTTTAAAGTCAACAAGATGTTACACCG 1200
QY 1201 GGGTCTGCTCTTGATAAATTAACGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCA 1260
Db 1201 GGGTCTGCTCTTGATAAATTAACGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCA 1260

QY 1261 AGAAGTGTGTGGCACCAGATGTCTTGTGCTGAAAAATGAAAGCTCAGAGAGACACCCAT 1320
Db 1261 AGAAGTGTGTGGCACCAGATGTCTTGTGCTGAAAAATGAAAGCTCAGAGAGACACCCAT 1320
QY 1321 CATTTGGTCAACTATATTTCCCGAGGTTCAATAGATTCACTCTTTGAAGGAAGTGTAC 1380
Db 1321 CATTTGGTCAACTATATTTCCCGAGGTTCAATAGATTCACTCTTTGAAGGAAGTGTAC 1380
QY 1381 TTAGTTAGGCTGGATGAAAAAGCAGAGAAGAACTTAAGCTCGGCGTCCCAATGAT 1440
Db 1381 TTAGTTAGGCTGGATGAAAAAGCAGAGAAGAACTTAAGCTCGGCGTCCCAATGAT 1440
QY 1441 GACACTTTGGATGAAGGAGTAGAGACTTGTGCAATTCAAACATTAAGCAACTGACATATCCA 1500
Db 1441 GACACTTTGGATGAAGGAGTAGAGACTTGTGCAATTCAAACATTAAGCAACTGACATATCCA 1500
QY 1501 AGCCCTGCCAAGAAAGTAAACCAAGACTCCCTGCCACAGCAGCAACCTGAAAGCACTGTC 1560
Db 1501 AGCCCTGCCAAGAAAGTAAACCAAGACTCCCTGCCACAGCAGCAACCTGAAAGCACTGTC 1560
QY 1561 ATTAGTAAATGGGGAACATTAAGATACCTGTGAGGTGCAAGACTTCAAGGTTGGGTGGGC 1620
Db 1561 ATTAGTAAATGGGGAACATTAAGATACCTGTGAGGTGCAAGACTTCAAGGTTGGGTGGGC 1620
QY 1621 ATGGGTTGGGGGTATGGGAACAAGTTGGAGGAATGGGATATCTGGGGATAATTTAAAGGA 1680
Db 1621 ATGGGTTGGGGGTATGGGAACAAGTTGGAGGAATGGGATATCTGGGGATAATTTAAAGGA 1680
QY 1681 TTACATGTTATGTAATTTTATGTGACTGACATGAGCGCTGGATGACTATGCTGTACTT 1740
Db 1681 TTACATGTTATGTAATTTTATGTGACTGACATGAGCGCTGGATGACTATGCTGTACTT 1740
QY 1741 GGGAAAGTCTTTGCTCTAATTTGCTGACATGCTTCTGTTGTGTGCTGGCCAATGCCAA 1800
Db 1741 GGGAAAGTCTTTGCTCTAATTTGCTGACATGCTTCTGTTGTGTGCTGGCCAATGCCAA 1800
QY 1801 ATGTACTGCAATGATGTTAAGGGCTCTGTAAAACTTCATACCTTTTGCCATTTGTATG 1860
Db 1801 ATGTACTGCAATGATGTTAAGGGCTCTGTAAAACTTCATACCTTTTGCCATTTGTATG 1860
QY 1861 CATGATGTTGGTTTTTAAACATGTAATAAGAAATGTAATGTGTACTTCTGTCAAGAAAGCA 1920
Db 1861 CATGATGTTGGTTTTTAAACATGTAATAAGAAATGTAATGTGTACTTCTGTCAAGAAAGCA 1920
QY 1921 GAGGTACTAATCTCCAAATTAATAATTTTAAACATGTAAATAAAAAAAAAAAAAAAAAA 1980
Db 1921 GAGGTACTAATCTCCAAATTAATAATTTTAAACATGTAAATAAAAAAAAAAAAAAAAAA 1980
QY 1981 AAAAAAAAAAAAAAAAAA 2002
Db 1981 AAAAAAAAAAAAAAAAAA 2002

RESULT 2
US-10-622-516-1
; Sequence 1, Application US/10622516
; Publication No. US20040018545A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001195DIV2
; CURRENT APPLICATION NUMBER: US/10/622, 516
; PRIOR FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 10/193, 295
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/819, 993
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2002

TYPE: DNA
ORGANISM: Human
us-10-622-516-1

Query Match 100.0%; Score 2002; DB 16; length 2002;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CGCCTCCAGCGACTCTCGGCACTGCCGGAGTCGGGTGGGTTGGCGGTATAAAGCTGCT	60
Db	1	CGCCTCCAGCGACTCTCGGCACTGCCGGAGTCGGGTGGGTTGGCGGTATAAAGCTGCT	60
QY	61	AGCGAAGGGGAGGCGCGCGGACTGTCCTTTCGTGGCTCACTCCCTTTCCTGCTGCGG	120
Db	61	AGCGAAGGGGAGGCGCGCGGACTGTCCTTTCGTGGCTCACTCCCTTTCCTGCTGCGG	120
QY	121	CTCGGTACGCTTGCTCTTTCACCATGCGCTGATCACTTCCTTGAATGCAGAACCTGTC	180
Db	121	CTCGGTACGCTTGCTCTTTCACCATGCGCTGATCACTTCCTTGAATGCAGAACCTGTC	180
QY	181	TGGCCAAAAGATGTGGAAATTGTGCCCTTGAGATCTAATTTCTTCTCAATATGTTGAT	240
Db	181	TGGCCAAAAGATGTGGAAATTGTGCCCTTGAGATCTAATTTCTTCTCAATATGTTGAT	240
QY	241	CAAGCAGAGTTGGAATAATGATGCTGTAGATGCTGGAAGTATACCATTTGGCTGGGC	300
Db	241	CAAGCAGAGTTGGAATAATGATGCTGTAGATGCTGGAAGTATACCATTTGGCTGGGC	300
QY	301	CAGGCCAAGATGGGCTTCTGCACAGATAGAGAAGATATTACTCTTTGCATGACTGTG	360
Db	301	CAGGCCAAGATGGGCTTCTGCACAGATAGAGAAGATATTACTCTTTGCATGACTGTG	360
QY	361	GTTTCAGAACTTATGAGAGAAATAACCTTTCCTATGATGCAATGGCGGCTGGAAGTT	420
Db	361	GTTTCAGAACTTATGAGAGAAATAACCTTTCCTATGATGCAATGGCGGCTGGAAGTT	420
QY	421	GGAACAGAGACAATCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGAGCTGTT	480
Db	421	GGAACAGAGACAATCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGAGCTGTT	480
QY	481	GAAAGTCTGGGAATACAGATATAGAAGAAATGCACACAATAATGCATGCTATGAGGC	540
Db	481	GAAAGTCTGGGAATACAGATATAGAAGAAATGCACACAATAATGCATGCTATGAGGC	540
QY	541	ACAGCTGCTGTTCTTCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCG	600
Db	541	ACAGCTGCTGTTCTTCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCG	600
QY	601	GGGACACATATGCACATGCTTAATGATTTTACAAAGCCTGATATGCTATCTGAATATCT	660
Db	601	GGGACACATATGCACATGCTTAATGATTTTACAAAGCCTGATATGCTATCTGAATATCT	660
QY	661	ATAGTAGATGAAAACTCTCCATACAGTGTCACTCAGTGCATTAGACCGCTGCTACTCT	720
Db	661	ATAGTAGATGAAAACTCTCCATACAGTGTCACTCAGTGCATTAGACCGCTGCTACTCT	720
QY	721	GTTCACTGCAAAAAGATCCATGCCAGTGGCAGAAAGGAAATGATAAAGATTTTACC	780
Db	721	GTTCACTGCAAAAAGATCCATGCCAGTGGCAGAAAGGAAATGATAAAGATTTTACC	780
QY	781	TTGAATGATTTTGGCTTCATGATCTTTCACTCACCATATTGTAACCTGGTTCAGAAATCT	840
Db	781	TTGAATGATTTTGGCTTCATGATCTTTCACTCACCATATTGTAACCTGGTTCAGAAATCT	840
QY	841	CTAGCTCGGATGTTGCTGAATGATCTCTTAATGACCAGAAATAGATTAATAATAGTATC	900
Db	841	CTAGCTCGGATGTTGCTGAATGATCTCTTAATGACCAGAAATAGATTAATAATAGTATC	900
QY	901	TATAGTGGCTGGAAGCCTTTGGGATGTTAAATTAGAAGACACTATTGATAGAGAT	960
Db	901	TATAGTGGCTGGAAGCCTTTGGGATGTTAAATTAGAAGACACTATTGATAGAGAT	960
QY	961	GTGGAAGAGCATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAAAGCATCT	1020

Db	961	GTGGAAGAGCATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAAAGCATCT	1020
QY	1021	TTACTTGTATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTGCA	1080
Db	1021	TTACTTGTATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTGCA	1080
QY	1081	TCTGTTAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAGAAATTGGAGTGTTCCT	1140
Db	1081	TCTGTTAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAGAAATTGGAGTGTTCCT	1140
QY	1141	TATGCTTGTGTTGGCTGCCACTCTGTACTCTTTAAAGTCAACAAGATGCTACACCG	1200
Db	1141	TATGCTTGTGTTGGCTGCCACTCTGTACTCTTTAAAGTCAACAAGATGCTACACCG	1200
QY	1201	GGGTCTGCTCTTGATAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTGATTC	1260
Db	1201	GGGTCTGCTCTTGATAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTGATTC	1260
QY	1261	AGAACTGCTGTCGCCACCATGCTCTTCCGTGAAAACATGAAGCTCAGAGAGACCCAT	1320
Db	1261	AGAACTGCTGTCGCCACCATGCTCTTCCGTGAAAACATGAAGCTCAGAGAGACCCAT	1320
QY	1321	CATTGGTCAACTATATTCGCCAGGTTCAATAGATTCACTCTTTGAAGAACGTGGTAC	1380
Db	1321	CATTGGTCAACTATATTCGCCAGGTTCAATAGATTCACTCTTTGAAGAACGTGGTAC	1380
QY	1381	TTAGTTAGGGTGATGAAAAGCACAGAAGACTTACGCTCGGCGTCCCACTCCAATGAT	1440
Db	1381	TTAGTTAGGGTGATGAAAAGCACAGAAGACTTACGCTCGGCGTCCCACTCCAATGAT	1440
QY	1441	GACACTTTGATGAAGAGTAGAAGTGTGTCATTGCAATGCAACTGAGCATATTCCA	1500
Db	1441	GACACTTTGATGAAGAGTAGAAGTGTGTCATTGCAATGCAACTGAGCATATTCCA	1500
QY	1501	AGCCCTGCCAAGAAATACCAAGACTCCCTGCCACAGACAGAACTGGAAGCAGCTGTC	1560
Db	1501	AGCCCTGCCAAGAAATACCAAGACTCCCTGCCACAGACAGAACTGGAAGCAGCTGTC	1560
QY	1561	ATTAATTAATGGGAAACATTAAGATTACTCTGTGAGGTGCAAGACTTCAGGGTGGGTC	1620
Db	1561	ATTAATTAATGGGAAACATTAAGATTACTCTGTGAGGTGCAAGACTTCAGGGTGGGTC	1620
QY	1621	ATGGGTGGGGTATGGGAACAGTTGGAGGAATGGATATCTGGGGATTAATTTTAAAGGA	1680
Db	1621	ATGGGTGGGGTATGGGAACAGTTGGAGGAATGGATATCTGGGGATTAATTTTAAAGGA	1680
QY	1681	TTACATGTTATGTAATTTTATGTGACTGACATGAGCCTGGATGACTATCGTGTACTT	1740
Db	1681	TTACATGTTATGTAATTTTATGTGACTGACATGAGCCTGGATGACTATCGTGTACTT	1740
QY	1741	GGGAAAGTCTCTTGTCTATTTGCTGACATGCTTCTGTGTGCTCTGGCCAATGCCAA	1800
Db	1741	GGGAAAGTCTCTTGTCTATTTGCTGACATGCTTCTGTGTGCTCTGGCCAATGCCAA	1800
QY	1801	ATGTACTGCAATGATGTTAAGGGCTCTGTAAACCTTCAATACCTCTTTGGCCAATTTGTA	1860
Db	1801	ATGTACTGCAATGATGTTAAGGGCTCTGTAAACCTTCAATACCTCTTTGGCCAATTTG	1860
QY	1861	CATGATGTTGTTTAAACATGCTATTAATGAATGTGTACTTCTGTCAAGAGAAAGCA	1920
Db	1861	CATGATGTTGTTTAAACATGCTATTAATGAATGTGTACTTCTGTCAAGAGAAAGCA	1920
QY	1921	GAGGTACTAATCTCCAATTAATAAATTTTAAACATGTAAATAAAAAAAAAAAAAA	1980
Db	1921	GAGGTACTAATCTCCAATTAATAAATTTTAAACATGTAAATAAAAAAAAAAAAAA	1980
QY	1981	AAAAAAAAAAAAAAAAAAAAA 2002	
Db	1981	AAAAAAAAAAAAAAAAAAAAA 2002	


```
US-09-919-039-98
; Sequence 98, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 98
; LENGTH: 3722
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 347055.4
;
; NAME/KEY: unsure
; LOCATION: 104, 2838
; OTHER INFORMATION: a, t, c, g, or other
US-09-919-039-98

Query Match      84.6%; Score 1693; DB 10; Length 3722;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1841; Conservative 0; Mismatches 20; Indels 126; Gaps 1;

QY      134 GCTCTTTCACCATGCGCTGCATCACTTCCTTGAATGCAGAGCTTGCGCCAAAAGATG 193
      |||||||
Db      238 GCTCTTTCACCATGCGCTGCATCACTTCCTTGAATGCAGAGCTTGCGCCAAAAGATG 297

QY      194 TGGGAATGTTGCCCTTGAGATCTATTTTCTCTCAATATGTTGATCAAGCAGAGTTGG 253
      |||||||
Db      298 TGGGAATGTTGCCCTTGAGATCTATTTTCTCTCAATATGTTGATCAAGCAGAGTTGG 357

QY      254 AAAAATATGATGGTGTAGATGCTGGAAGATATACCATTTGGCTTGGCCAGGCCAAGATGG 313
      |||||||
Db      358 AAAAATATGATGGTGTAGATGCTGGAAGATATACCATTTGGCTTGGCCAGGCCAAGATGG 417

QY      314 GCTTCTGCACAGATAGAGAATATTTAACTCTTTGCATGACTGTGTTCAAGATCTTA 373
      |||||||
Db      418 GCTTCTGCACAGATAGAGAATATTTAACTCTTTGCATGACTGTGTTCAAGATCTTA 477

QY      374 TGGAGAGAAAATAACCTTTCCTATGATTGCATGCGGCGCTGGAAGTTGGAACAGACAA 433
      |||||||
Db      478 TGGAGAGAAAATAACCTTTCCTATGATTGCATGCGGCGCTGGAAGTTGGAACAGACAA 537

QY      434 TCATGCACAAATCAAAGCTGTGGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGA 493
      |||||||
Db      538 TCATGCACAAATCAAAGCTGTGGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGA 597

QY      494 ATACAGATATAGAAGGAATCGACACAATAATGCTATGGAAGGACACAGCTGCTGCT 553
      |||||||
Db      598 ATACAGATATAGAAGGAATCGACACAATAATGCTATGGAAGGACACAGCTGCTGCT 657

QY      554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT----- 591
      |||||||
Db      658 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGAAGGATATGCCCTGTAAGTTG 717

QY      592 ----- 591
Db      718 CAGGAGATATGCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGACAG 777

QY      592 -----GGGCTTCGTGGGACAC 607
Db      778 TAGCTCTGCTAATTGGGCCAAATGCTCTTTAATTTTGAACGAGGGCTTCGTGGGACAC 837

QY      608 ATATGCAACATGCTATGATTTTAAAGCCCTGATATGCTATCTGAATATCCTATAGTAG 667
      |||||||
Db      838 ATATGCAACATGCTATGATTTTAAAGCCCTGATATGCTATCTGAATATCCTATAGTAG 897
```

```
QY      668 ATGAAAACCTCTCCATACAGTGTCTACCTGACGTGATTAAGCCGCTGCTACTCTGTCTACT 727
      |||||||
Db      898 ATGAAAACCTCTCCATACAGTGTCTACCTGACGTGATTAAGCCGCTGCTACTCTGTCTACT 957

QY      728 GCAAAAAGATCCATGCCCGCAGTGGCAGAAAGGGAAATGATTAAGATTTTACCTGAATG 787
      |||||||
Db      958 GCAAAAAGATCCATGCCCGCAGTGGCAGAAAGGGAAATGATTAAGATTTTACCTGAATG 1017

QY      788 ATTTGGCTTCATGATCTTTCACTCACTCAACCATATTTGTAACCTGCTCAGAAATCTAGCTC 847
      |||||||
Db      1018 ATTTGGCTTCATGATCTTTCACTCACTCAACCATATTTGTAACCTGCTCAGAAATCTAGCTC 1077

QY      848 GGATGTTGCTGAATGACTTCCTTAATGACCAAGATAGAGATTAATAATAGTATCTATAGTG 907
      |||||||
Db      1078 GGATGTTGCTGAATGACTTCCTTAATGACCAAGATAGAGATTAATAATAGTATCTATAGTG 1137

QY      908 GCCTGGAAGCCTTTGGGAGTGTAAATTAGAAACACCTTACTTTGATAGAGATGGAGAGA 967
      |||||||
Db      1138 GCCTGGAAGCCTTTGGGAGTGTAAATTAGAAACACCTTACTTTGATAGAGATGGAGAGA 1197

QY      968 AGGCATTTATGAAGGCTAGCTCTGAACCTCTCAGTCAGAAAACAAAGGCATCTTACTTG 1027
      |||||||
Db      1198 AGGCATTTATGAAGGCTAGCTCTGAACCTCTCAGTCAGAAAACAAAGGCATCTTACTTG 1257

QY      1028 TATCAAAATCAAAATGAAAATATGTACACATCTTCAGTATATGTTCCCTGCACTCTGTTTC 1087
      |||||||
Db      1258 TATCAAAATCAAAATGAAAATATGTACACATCTTCAGTATATGTTCCCTGCACTCTGTTTC 1317

QY      1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAAATGAGTGTGTTTCTTATGGTT 1147
      |||||||
Db      1318 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAAATGAGTGTGTTTCTTATGGTT 1377

QY      1148 CTGCTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACACAGATGCTACACCGGGCTCTG 1207
      |||||||
Db      1378 CTGCTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACACAGATGCTACACCGGGCTCTG 1437

QY      1208 CTCTTGATTAATAATACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTTCAAGAACTG 1267
      |||||||
Db      1438 CTCTTGATTAATAATACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTTCAAGAACTG 1497

QY      1268 GTGTGGCACACAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATGATTTGG 1327
      |||||||
Db      1498 GTGTGGCACACAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATGATTTGG 1557

QY      1328 TCAACTATATTTCCCGAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGTAAGTTAAGTTA 1387
      |||||||
Db      1558 TCAACTATATTTCCCGAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGTAAGTTAAGTTA 1617

QY      1388 GCGTGGATGAAAAGCACAGAGAAGACTTACGCTCGGCGTCCCACTCCAATGATGACACTT 1447
      |||||||
Db      1618 GCGTGGATGAAAAGCACAGAGAAGACTTACGCTCGGCGTCCCACTCCAATGATGACACTT 1677

QY      1448 TGGATGAAGGATAGGACTTGTGCAATTAAGCATAGCAACTGACATATTTCCAAGCCCTG 1507
      |||||||
Db      1678 TGGATGAAGGATAGGACTTGTGCAATTAAGCATAGCAACTGACATATTTCCAAGCCCTG 1737

QY      1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAAGAACTGAAAGCAGCTGTCAATTAAGTA 1567
      |||||||
Db      1738 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAAGAACTGAAAGCAGCTGTCAATTAAGTA 1797

QY      1568 ATGGGGAACATTAAAGATCTCTGTGAGGTGCAAGACTTCAAGGTTGGGTTGGGATGGGGT 1627
      |||||||
Db      1798 ATGGGGAACATTAAAGATCTCTGTGAGGTGCAAGACTTCAAGGTTGGGTTGGGATGGGGT 1857

QY      1628 GGGGCTATGGGAACAGTTGGAAGAAATGGGATATCTGGGGATTAATTTAAAGGATTAACATG 1687
      |||||||
Db      1858 GGGGCTATGGGAACAGTTGGAAGAAATGGGATATCTGGGGATTAATTTAAAGGATTAACATG 1917

QY      1688 TTATGTAATTTTATGTGATGACATGAGCCCTGGATGACATATCGTGTACTTGGGAAAG 1747
      |||||||
Db      1918 TTATGTAATTTTATGTGATGACATGAGCCCTGGATGACATATCGTGTACTTGGGAAAG 1977
```


QY	1208	CTCTTGATAAATTAACAGCAAGTTATATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTG	1267
Db	1211	CTCTTGATAAATTAACAGCAAGTTATATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTG	1270
QY	1268	GTGTGGCACCAGATGTCTTCGCTGAAACATGAAGCTCAGAGAGACACCCATCATTTGG	1327
Db	1271	GTGTGGCACCAGATGTCTTCGCTGAAACATGAAGCTCAGAGAGACACCCATCATTTGG	1330
QY	1328	TCAACTATATTCGCCAGGGTTCAATAGATTCACTCTTGAAGAACGTGTACTTAGTTA	1387
Db	1331	TCAACTATATTCGCCAGGGTTCAATAGATTCACTCTTGAAGAACGTGTACTTAGTTA	1390
QY	1388	GGGTGATGAAAAGCACAGAGAAGTAAGTACGCTCGGCGTCCCACTCCAATGATGACACTT	1447
Db	1391	GGGTGATGAAAAGCACAGAGAAGTAAGTACGCTCGGCGTCCCACTCCAATGATGACACTT	1450
QY	1448	TGGATGAAGAGTAGGACTTGTGCAATTCAAACATATGCAACTGAGCATATTTCCAAGCCCTG	1507
Db	1451	TGGATGAAGAGTAGGACTTGTGCAATTCAAACATATGCAACTGAGCATATTTCCAAGCCCTG	1510
QY	1508	CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACTGAAAGCAGCTGCTATTAGTA	1567
Db	1511	CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACTGAAAGCAGCTGCTATTAGTA	1570
QY	1568	ATGGGGAAACATTAAAGATACTCTGTGAGGCTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT	1627
Db	1571	ATGGGGAAACATTAAAGATACTCTGTGAGGCTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT	1630
QY	1628	GGGGGTATGGGAACAAGTTGG	1647
Db	1631	GGGGGTATGGGAACAAGTTGG	1650

```

RESULT 5
US-10-236-417-183
; Sequence 183, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
;   APPLICANT: Agee et al.
;   TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;   FILE REFERENCE: 21402-442C
;   CURRENT APPLICATION NUMBER: US/10/236,417
;   CURRENT FILING DATE: 2003-01-06
;   PRIOR APPLICATION NUMBER: US60/318,120
;   PRIOR FILING DATE: 2001-09-01
;   PRIOR APPLICATION NUMBER: US60/318,430
;   PRIOR FILING DATE: 2001-09-10
;   PRIOR APPLICATION NUMBER: US60/322,781
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/318,184
;   PRIOR FILING DATE: 2001-09-07
;   PRIOR APPLICATION NUMBER: US60/361,663
;   PRIOR FILING DATE: 2002-03-05
;   PRIOR APPLICATION NUMBER: US60/396,412
;   PRIOR FILING DATE: 2002-07-17
;   PRIOR APPLICATION NUMBER: US60/322,636
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/322,817
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/322,816
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/323,519
;   PRIOR FILING DATE: 2001-09-19
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 341
;   SOFTWARE: Custom
;   SEQ ID NO 183
;   LENGTH: 1650
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;

```

```

; LOCATION: (22) .. (1582)
US-10-236-417-183

Query Match          68.4%; Score 1370; DB 13; Length 1650;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1.

```

QY	194	TGGCAATGTTGCCCTGAGATCTAATTTCTCTCTCAAATATGTTGATCAAGACAGAGTTGG	253
Db	71	TTGGGATTTGTCCTTGAGATCTAATTTCTCTCTCAAATATGTTGATCAAGACAGAGTTGG	130
QY	254	AAAAATATGATGGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGG	313
Db	131	AAAAATATGATGGTGTAGATGCTGGGAAGTATACCATTGGCTTGGGCCAGGCCAAGATGG	190
QY	314	GCTTCTGCACAGATAGAGAAATATTAATCTCTCTTTCATGACTGTGGTTCAATCTTA	373
Db	191	GCTTCTGCACAGATAGAGAAATATTAATCTCTCTTTCATGACTGTGGTTCAATCTTA	250
QY	374	TGGAGAGAAATTAACCTTTCTTATGATTTGCCATTGGCGCGCTGGAAGTTGGAACAGACAA	433
Db	251	TGGAGAGAAATTAACCTTTCTTATGATTTGGCGCGCTGGAAGTTGGAACAGACAA	310
QY	434	TCATCCACAATCAAGTCTGTGAAGACTAATTTGATGACAGCTGTTGAAGAGTCTGGGA	493
Db	311	TCATCCACAATCAAGTCTGTGAAGACTAATTTGATGACAGCTGTTGAAGAGTCTGGGA	370
QY	494	ATACAGATATAGAAGGATCGACACAACAATAATGCATGCTATGGAAGCACAGCTGCTCT	553
Db	371	ATACAGATATAGAAGGATCGACACAACAATAATGCATGCTATGGAAGCACAGCTGCTCT	430
QY	554	TCAATGCTGTTAACTGAGTGTAGTCCAGCTCTTGGGAT	591
Db	431	TCAATGCTGTTAACTGAGTGTAGTCCAGCTCTTGGGATGACGGTATGCCCTGTAGTTG	490
QY	592		591
Db	491	CAGAGATATTGCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGACAG	550
QY	592		591
Db	551	TAGCTCTGTAATTGGGCCAATGCTCTTTAATTTTGAACGAGGGCTTCGTGGACAC	610
QY	608	ATATGCAACATGCTTATGATTTTACAGCGCTGATATGCTATCTGAATATCTCTATAGTAG	667
Db	611	ATATGCAACATGCTTATGATTTTACAGCGCTGATATGCTATCTGAATATCTCTATAGTAG	670
QY	668	ATGGAACCTCTCCATACAGTGTCTACTCAATGATTAAGCCGCTGCTACTCTGTACT	727
Db	671	ATGGAACCTCTCCATACAGTGTCTACTCAATGATTAAGCCGCTGCTACTCTGTACT	730
QY	728	GCAAAAAGATCCATGCCCAAGTGGCAGAAAGAGGAAATGATAAAGATTTTACCTTGAATG	787
Db	731	GCAAAAAGATCCATGCCCAAGTGGCAGAAAGAGGAAATGATAAAGATTTTACCTTGAATG	790
QY	788	ATTTGGCTTCATGATCTTTCACTCAACCATATGTAACCTGTTCAAGAACTCTAGCTC	847
Db	791	ATTTGGCTTCATGATCTTTCACTCAACCATATGTAACCTGTTCAAGAACTCTAGCTC	850
QY	848	GGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTG	907
Db	851	GGATGTTGCTGAATGACTTCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTG	910
QY	908	GCCTGAAGCCTTTGGGATGTTAAATTAGAAGACACTACTTGTGATAGAGATGTGAGA	967
Db	911	GCCTGAAGCCTTTGGGATGTTAAATTAGAAGACACTACTTGTGATAGAGATGTGAGA	970
QY	968	AGGCATTATGAAGGCTAGCTGTGAACCTCTTCACTCAGAAAACAAAGCATCTTACTTG	1027


```

Db      971 AGGCATTATGAAGGCTAGCTTGAACCTCTTCAGTCAGAAAACAAAAGCATCTTACTTG 1030
QY      1028 TATCAATCAAAATGGAATATGTACACATCTTCAGTATATGGTCCCTTGCATCTGTC 1087
Db      1031 TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGGTCCCTTGCATCTGTC 1090
QY      1088 TAGCACAGTACTCACCCTCAGCAATTAGCAGGGAAGAAATGAGTGTCTTCTTATGTT 1147
Db      1091 TAGCACAGTACTCACCCTCAGCAATTAGCAGGGAAGAAATGAGTGTCTTCTTATGTT 1150
QY      1148 CTGGTTTGGCTGCCACTCTGTACTCTCTTAAGTCACACAGATGCTACACCGGGTCTG 1207
Db      1151 CTGGTTTGGCTGCCACTCTGTACTCTCTTAAGTCACACAGATGCTACACCGGGTCTG 1210
QY      1208 CTCTTGATAAATACAGCAAGTTATGTGATCTTAATCAAGGCTTGATTCAGAACTG 1267
Db      1211 CTCTTGATAAATACAGCAAGTTATGTGATCTTAATCAAGGCTTGATTCAGAACTG 1270
QY      1268 GTGTGGCACCAGATGTCCTTGGTGAAAACATGAAGCTCAGAGAGACACCCATCATTTG 1327
Db      1271 GTGTGGCACCAGATGTCCTTGGTGAAAACATGAAGCTCAGAGAGACACCCATCATTTG 1330
QY      1328 TCAACTATATTTCCCGAGGGTCAATAGATTCACTCTTTGAAGAACTGTGACTTAGTTA 1387
Db      1331 TCAACTATATTTCCCGAGGGTCAATAGATTCACTCTTTGAAGAACTGTGACTTAGTTA 1390
QY      1388 GGGTGGATGAAAGCACAGAAAGACTTAAGCTCGGCTGCCACTCCAATGATGACACTT 1447
Db      1391 GGGTGGATGAAAGCACAGAAAGACTTAAGCTCGGCTGCCACTCCAATGATGACACTT 1450
QY      1448 TGGATGAAGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTTCCAAGCCCTG 1507
Db      1451 TGGATGAAGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTTCCAAGCCCTG 1510
QY      1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACTGGAAGCAGCTGTCATTAGTA 1567
Db      1511 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACTGGAAGCAGCTGTCATTAGTA 1570
QY      1568 ATGGGGAACATTAAAGTACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGT 1627
Db      1571 ATGGGGAACATTAAAGTACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGT 1630
QY      1628 GGGGGTATGGGAACAGTTGG 1647
Db      1631 GGGGGTATGGGAACAGTTGG 1650

```

```

RESULT 6
US-10-236-417-185
; Sequence 185, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17

```

```

; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 185
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22) .. (1582)
US-10-236-417-185

```

```

Query Match      68.4%; Score 1370; DB 13; Length 1650;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

```

```

QY      134 GCTCTTCAACCATGCGCTGGAATCACTTCTTGAATGCAAGCTTGCGCCAAAAGATG 193
Db      11  GCTCTTCAACCATGCGCTGGAATCACTTCTTGAATGCAAGCTTGCGCCAAAAGATG 70
QY      194 TGGGAATTGTTGCCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGAGTTGG 253
Db      71  TTGGATTTGTTGCCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGAGTTGG 130
QY      254 AAAAATATGATGTTGATGATGCTGGAAGATATACCATTTGGCTTGGGCCAAGATGG 313
Db      131 AAAAATATGATGTTGATGATGCTGGAAGATATACCATTTGGCTTGGGCCAAGATGG 190
QY      314 GCTTCTGCACAGATAGAGAATATTTAATCTCTTTGCAATGACTGTGTTCAGAATCTTA 373
Db      191 GCTTCTGCACAGATAGAGAATATTTAATCTCTTTGCAATGACTGTGTTCAGAATCTTA 250
QY      374 TGGAGAAATTAACCTTCTATGATTGTCATTGGCGCGCTGGAAGTTGGAACAGAGCAA 433
Db      251 TGGAGAAATTAACCTTCTATGATTGTCATTGGCGCGCTGGAAGTTGGAACAGAGCAA 310
QY      434 TCATCGACAAATCAAGTCTGTGAAGCTAATTTGATGCACTGTTTGAAGATCTGGGA 493
Db      311 TCATCGACAAATCAAGTCTGTGAAGCTAATTTGATGCACTGTTTGAAGATCTGGGA 370
QY      494 ATACAGATATAGAAGGAATGCACACAATAATGCATGCTATGGAAGCACAGCTGCTGCT 553
Db      371 ATACAGATATAGAAGGAATGCACACAATAATGCATGCTATGGAAGCACAGCTGCTGCT 430
QY      554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT----- 591
Db      431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGACGGTATGCCCTGTAGTTG 490
QY      592 ----- 591
Db      491 CAGAGATATTGCTGTATATGCCACAGAAATGCTAGACTTACAGGTGAGTTGAGCAG 550
QY      592 -----GGGCTTGTGGGAC 607
Db      551 TAGCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTGTGGGAC 610
QY      608 ATATGCAACATGCGCTATGATTTTTCACAGCGCTGATATGCTATCTGAATATCCTATAGTAG 667
Db      611 ATATGCAACATGCGCTATGATTTTTCACAGCGCTGATATGCTATCTGAATATCCTATAGTAG 670
QY      668 ATGAAAACTCTCCATACAGTGTACTACCTCAGTGCAATTAGACCGCTGCTACTGTGTACT 727
Db      671 ATGGGAACTCTCCATACAGTGTACTACCTCAGTGCAATTAGACCGCTGCTACTGTGTACT 730
QY      728 GCAAAAAAGATCCATGCCAGTGGCAGAAAAGAGGGAATGATTAAGATTTCCTTGAATG 787
Db      731 GCAAAAAAGATCCATGCCAGTGGCAGAAAAGAGGGAATGATTAAGATTTCCTTGAATG 790
QY      788 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAACTGGTTCAGAAATCTCTAGCTC 847

```

|||||
Db 791 ATTTGGCTTCATGATCTTTCACTCACCATATTGTAACGTGGTTCAGAAATCTTAGCTC 850
QY 848 GGATGTTGCTGAATGACTTCTTAATGACCAAGATAGATAAAAAATGATCTATAGTG 907
Db 851 GGATGTTGCTGAATGACTTCTTAATGACCAAGATAGATAAAAAATGATCTATAGTG 910
QY 908 GCGTGAAGCCTTTGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA 967
Db 911 GCGTGAAGCCTTTGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA 970
QY 968 AGGCATTATGAAAGGCTAGCTCTGAACCTCTCAGTCAGAAAACAAAGGCATCTTACTTG 1027
Db 971 AGGCATTATGAAAGGCTAGCTCTGAACCTCTCAGTCAGAAAACAAAGGCATCTTACTTG 1030
QY 1028 TATCAAAATCAAAATGMAAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTC 1087
Db 1031 TATCAAAATCAAAATGMAAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTC 1090
QY 1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTCTTATGTT 1147
Db 1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTCTTATGTT 1150
QY 1148 CTGTTGGCTGCGCACTCTGTACTCTTAAAGTCACACAGATGCTACACGGGCTG 1207
Db 1151 CTGTTGGCTGCGCACTCTGTACTCTTAAAGTCACACAGATGCTACACGGGCTG 1210
QY 1208 CTCTGATAAAATTAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATCAAGAACTG 1267
Db 1211 CTCTGATAAAATTAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATCAAGAACTG 1270
QY 1268 GTGTGGCACCAGATGTCTTGGCTGMAAACAATGAAGCTCAGAGAGGACACCATCATTTGG 1327
Db 1271 GTGTGGCACCAGATGTCTTGGCTGMAAACAATGAAGCTCAGAGAGGACACCATCATTTGG 1330
QY 1328 TCAACTATATTTCCCGAGGGTTCAATAGATTCACTTTGMAAGGAACGTGTACTTAGTTA 1387
Db 1331 TCAACTATATTTCCCGAGGGTTCAATAGATTCACTTTGMAAGGAACGTGTACTTAGTTA 1390
QY 1388 GGGTGATGAAAAAGCACAGAAACTTACCGTCGGCGTCCCACTCCAATGATGACACTT 1447
Db 1391 GGGTGATGAAAAAGCACAGAAACTTACCGTCGGCGTCCCACTCCAATGATGACACTT 1450
QY 1448 TGGATGAAGAGTAGGACTGTGTCATTCAACATAGCAACTGAGCATATTCCAAGCCCTG 1507
Db 1451 TGGATGAAGAGTAGGACTGTGTCATTCAACATAGCAACTGAGCATATTCCAAGCCCTG 1510
QY 1508 CCAAGAAAGTACCAGACTCCCTGCCACAGAGCAAGAACTGAAGCAGCTGTCTTAGTA 1567
Db 1511 CCAAGAAAGTACCAGACTCCCTGCCACAGAGCAAGAACTGAAGCAGCTGTCTTAGTA 1570
QY 1568 ATGGGAACATTAAAGATCTGTGAGGTGCAAGCTTCAGGGTGGGTGGCATGGGGT 1627
Db 1571 ATGGGAACATTAAAGATCTGTGAGGTGCAAGCTTCAGGGTGGGTGGCATGGGGT 1630
QY 1628 GGGGTATGGGAACAGTTGG 1647
Db 1631 GGGGTATGGGAACAGTTGG 1650

RESULT 7
US-10-236-417-189
; Sequence 189, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430

; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 189
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1582)
US-10-236-417-189

Query Match 68.4%; Score 1370; DB 13; Length 1650;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;
QY 134 GCTCTTTCACCATGCGCTGATCACTTCCCTTGAATGACAGAGCTTGCGCAAAAGATG 193
Db 11 GCTCTTTCACCATGCGCTGATCACTTCCCTTGAATGACAGAGCTTGCGCAAAAGATG 70
QY 194 TGGGAATGTTGCCCTTGAGATCTATTTTCTCTCAATATGTGATCAAGCAGATTGG 253
Db 71 TTGGATGTGTCCTTGAGATCTATTTCTCTCAATATGTGATCAAGCAGATTGG 130
QY 254 AAAAATATGATGGTGTAGATGCTGMAAGTATACCATTTGGCTTGGGCCAGGCCAAGATTGG 313
Db 131 AAAAATATGATGGTGTAGATGCTGMAAGTATACCATTTGGCTTGGGCCAGGCCAAGATTGG 190
QY 314 GCTTTCGACAGATAGAGAGATATTAACTCTCTTTGCATGACTGTGTTCAGAACTTTA 373
Db 191 GCTTTCGACAGATAGAGAGATATTAACTCTCTTTGCATGACTGTGTTCAGAACTTTA 250
QY 374 TGGAGAGAAATTAACCTTTCTATGATTTGGCGGCTGGAAGTTGGAACAGACAA 433
Db 251 TGGAGAGAAATTAACCTTTCTATGATTTGGCGGCTGGAAGTTGGAACAGACAA 310
QY 434 TCATGACAAATCAAAAGTCTGTGAAGACTAATTGATGCAAGCTGTTGAAGAGTCTGGGA 493
Db 311 TCATGACAAATCAAAAGTCTGTGAAGACTAATTGATGCAAGCTGTTGAAGAGTCTGGGA 370
QY 494 ATACAGATATAGAAAGAAATGACACAACTAATGCATGCTATGAGGACACAGCTGCTGCT 553
Db 371 ATACAGATATAGAAAGAAATGACACAACTAATGCATGCTATGAGGACACAGCTGCTGCT 430
QY 554 TCAATGCTGTTAACTGATTTGATGCCAGCTCTTGGAT----- 591
Db 431 TCAATGCTGTTAACTGATTTGATGCCAGCTCTTGGATGACGGTATGCCCTGTAGTTG 490
QY 592 ----- 591
Db 491 CAGAGATATTGCTGTATATGCCACAGAAATGCTAGACCTACAGCTGAGTTGAGACAG 550
QY 592 -----GGGCTTGTGGGACAC 607
Db 551 TAGCTCTGCTAATTGGCCAAATGCTCCTTTAAATTTTGAACGAGGGCTTGTGGGACAC 610

QY

608

ATATGCAACATGCCCTATGATTTTAAAGCCCTGATATGCTATCTGAATATCCTATAGTAG

667

Db

611

ATATGCACATGCCCTATGATTTTAAAGCCCTGATATGCTATCTGAATATCCTATAGTAG

670

QY

668

ATGMAAACTCTCCATACAGTGGCTACTCAGTGCATTAGACCGGCTACTCTGTCTACT

727

Db

671

ATGGGAAACTCTCCATACAGTGGCTACTCAGTGCATTAGACCGGCTACTCTGTCTACT

730

QY

728

GCAAAAAGATCCATGCCAGTGGCAGAAAGAGGAAATGATTAAGATTTTACCCTGAATG

787

Db

731

GCAAAAAGATCCATGCCAGTGGCAGAAAGAGGAAATGATTAAGATTTTACCCTGAATG

790

QY

788

ATTTGGCTTCATGATCTTTCACTGACCATATTTGTAACCTGGTTCAGAAATCTTAGCTC

847

Db

791

ATTTGGCTTCATGATCTTTCACTGACCATATTTGTAACCTGGTTCAGAAATCTTAGCTC

850

QY

848

GGATGTTGCTGAATGACTTCCCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTG

907

Db

851

GGATGTTGCTGAATGACTTCCCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTG

910

QY

908

GCCGTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCCTACTTTGATAGAGATGTGAGA

967

Db

911

GCCGTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCCTACTTTGATAGAGATGTGAGA

970

QY

968

AGGCATTTATGAAGGCTAGCTCTGAATCTTCAGTCAAGAAACAAAGCATCTTTACTTG

1027

Db

971

AGGCATTTATGAAGGCTAGCTCTGAATCTTCAGTCAAGAAACAAAGCATCTTTACTTG

1030

QY

1028

TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTGCATCTGTTTC

1087

Db

1031

TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTGCATCTGTTTC

1090

QY

1088

TAGCACAGTACTCACCCTCAGCAATTACAGGGAAGAGATTTGGAGTGTTCCTATGCTT

1147

Db

1091

TAGCACAGTACTCACCCTCAGCAATTACAGGGAAGAGATTTGGAGTGTTCCTATGCTT

1150

QY

1148

CTGGTTTGGCTGCCACTCTGTACTCTTAAAGTCACACAGAAGTCTACACGGGGCTTG

1207

Db

1151

CTGGTTTGGCTGCCACTCTGTACTCTTAAAGTCACACAGAAGTCTACACGGGGCTTG

1210

QY

1208

CTCTTGATAAATTAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTG

1267

Db

1211

CTCTTGATAAATTAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTG

1270

QY

1268

GTTGGCACCATGATGTCTTCGCTGAAACATGAAGCTCAGAGAGACACCCATCATTTGG

1327

Db

1271

GTTGGCACCATGATGTCTTCGCTGAAACATGAAGCTCAGAGAGACACCCATCATTTGG

1330

QY

1328

TCAACTATATTCCCCAGGGTTCATAGATTCACTCTTTGAAGGAACTGGTACTTAGTTA

1387

Db

1331

TCAACTATATTCCCCAGGGTTCATAGATTCACTCTTTGAAGGAACTGGTACTTAGTTA

1390

QY

1388

GGGTGGATGAAAAAGCACAGAAGAACTTACGCTCGGCTGCCACTCCAATGATGACACTT

1447

Db

1391

GGGTGGATGAAAAAGCACAGAAGAACTTACGCTCGGCTGCCACTCCAATGATGACACTT

1450

QY

1448

TGGATGAAGGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTCCAAGCCCTG

1507

Db

1451

TGGATGAAGGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTCCAAGCCCTG

1510

QY

1508

CCAAGAAAGTACCAAGACTCCCTGCAACAGCAGACAGAACTGAAAGCAGCTGTCAATTAGTA

1567

Db

1511

CCAAGAAAGTACCAAGACTCCCTGCAACAGCAGACAGAACTGAAAGCAGCTGTCAATTAGTA

1570

QY

1568

ATGGGGAACATTAAAGATCTGTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT

1627

Db

1571

ATGGGGAACATTAAAGATCTGTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGT

1630

QY

1628

GGGGGTATGGGAACAGTTGG 1647

Db

1631

GGGGGTATGGGAACAGTTGG 1650

RESULT 8

US-10-236-417-193

Sequence 193, Application US/10236417

Publication No. US20040048256A1

GENERAL INFORMATION:

APPLICANT: Agee et al.

TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

FILE REFERENCE: 21402-442C

CURRENT APPLICATION NUMBER: US/10/236,417

PRIOR FILING DATE: 2003-01-06

PRIOR APPLICATION NUMBER: US60/318,120

PRIOR FILING DATE: 2001-09-01

PRIOR APPLICATION NUMBER: US60/318,430

PRIOR FILING DATE: 2001-09-10

PRIOR APPLICATION NUMBER: US60/322,781

PRIOR FILING DATE: 2001-09-17

PRIOR APPLICATION NUMBER: US60/318,184

PRIOR FILING DATE: 2001-09-07

PRIOR APPLICATION NUMBER: US60/361,663

PRIOR FILING DATE: 2002-03-05

PRIOR APPLICATION NUMBER: US60/396,412

PRIOR FILING DATE: 2002-07-17

PRIOR APPLICATION NUMBER: US60/322,636

PRIOR FILING DATE: 2001-09-17

PRIOR APPLICATION NUMBER: US60/322,817

PRIOR FILING DATE: 2001-09-17

PRIOR APPLICATION NUMBER: US60/322,816

PRIOR FILING DATE: 2001-09-17

PRIOR APPLICATION NUMBER: US60/323,519

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 341

SOFTWARE: Custom

SEQ ID NO 193

LENGTH: 1650

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (22)..(1582)

US-10-236-417-193

Query Match

Best Local Similarity 68.4%; Score 1370; DB 13; length 1650;

Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

QY

134

GCTCTTTCACCATGCGCTGATCACTTCCTTGAAATGCAGAACTTGCTGGCCAAAGATG

193

Db

11

GCTCTTTCACCATGCGCTGATCACTTCCTTGAAATGCAGAACTTGCTGGCCAAAGATG

70

QY

194

TGGGAATTGTTGCCCTTGAGATCTATTTTCCCTTCAATATGTGATCAAGCAGAGTTGG

253

Db

71

TTGGAATTGTTGCCCTTGAGATCTATTTTCCCTTCAATATGTGATCAAGCAGAGTTGG

130

QY

254

AAAAATATGATGTTAGATGCTGGAAGTATACCATGGCTTGGCCAGGCCAAGATGG

313

Db

131

AAAAATATGATGTTAGATGCTGGAAGTATACCATGGCTTGGCCAGGCCAAGATGG

190

QY

314

GCTTCTGCACAGATAGAGAGATATTAATCTCTTTGCATGACTGTGTTCAAGATCTTA

373

Db

191

GCTTCTGCACAGATAGAGAGATATTAATCTCTTTGCATGACTGTGTTCAAGATCTTA

250

QY

374

TGGAGAGAAATTAACCTTTCCTATGATTGCAATGGGCGGCTGGAAGTTGGAACAGACAA

433

Db

251

TGGAGAGAAATTAACCTTTCCTATGATTGCAATGGGCGGCTGGAAGTTGGAACAGACAA

310

QY

434

TCAATGACAAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGA

493

Db

311

TCAATGACAAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGA

370

QY

494

ATACAGATATAGGAAGAAATGCACAACTAATGCAATGCTATGAGGACACAGCTGCTGCT

553

Db

371

ATACAGATATAGGAAGAAATGCACAACTAATGCAATGCTATGAGGACACAGCTGCTGCT

430

QY	554	TCATGCTGTAACTGGATTGAGTCCAGCTCTGGGAT-----	591
Db	431	TCATGCTGTAACTGGATTGAGTCCAGCTCTGGGATGAGCGTATGCCCTGTAGTTG	490
QY	592	-----	591
Db	491	CAGAGATATTTCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGAGCAG	550
QY	592	-----GGGCTTGTGGGACAC	607
Db	551	TAGCTTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTGTGGACAC	610
QY	608	ATATGCAACATGCCCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG	667
Db	611	ATATGCAACATGCCCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG	670
QY	668	ATGAAAACCTCCCATACAGGTCTACCTCAGTGCATTAGACCGCTGCTACTGTCTACT	727
Db	671	ATGGAAAACCTCCCATACAGGTCTACCTCAGTGCATTAGACCGCTGCTACTGTCTACT	730
QY	728	GCAAAAAGATCCATGCCCAGTGGCAGAAAGGGAATGATAAAGATTTTACCTTGAATG	787
Db	731	GCAAAAAGATCCATGCCCAGTGGCAGAAAGGGAATGATAAAGATTTTACCTTGAATG	790
QY	788	ATTTGGCTTCATGATCTTTCACCTCACCATATTGTAACTGGTTCAGAAATCTTAGCTC	847
Db	791	ATTTGGCTTCATGATCTTTCACCTCACCATATTGTAACTGGTTCAGAAATCTTAGCTC	850
QY	848	GGATGTTGCTGAATGACTTCTTAAATGAGCAGAAATAGATATAATATCTATAGTG	907
Db	851	GGATGTTGCTGAATGACTTCTTAAATGAGCAGAAATAGATATAATATCTATAGTG	910
QY	908	GCCTGGAAGCCTTTGGGGATGTTAAATTAGAAACACCTACTTTGATAGAGATGTGAGA	967
Db	911	GCCTGGAAGCCTTTGGGGATGTTAAATTAGAAACACCTACTTTGATAGAGATGTGAGA	970
QY	968	AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCACTCAGAAAACAAAGGCATCTTACTTG	1027
Db	971	AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCACTCAGAAAACAAAGGCATCTTACTTG	1030
QY	1028	TATCAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTGACCTGTTTC	1087
Db	1031	TATCAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTGACCTGTTTC	1090
QY	1088	TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAAATGGATGTTTTCTTATGGTT	1147
Db	1091	TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAAATGGATGTTTTCTTATGGTT	1150
QY	1148	CTGGTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAAGATGTACACCGGGGTCTG	1207
Db	1151	CTGGTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAAGATGTACACCGGGGTCTG	1210
QY	1208	CTCTTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG	1267
Db	1211	CTCTTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG	1270
QY	1268	GTGTGGCACCAAGATGTCCTCGTGAATAACATGAAGCTCAGAGAGGACACCCCATATTGG	1327
Db	1271	GTGTGGCACCAAGATGTCCTCGTGAATAACATGAAGCTCAGAGAGGACACCCCATATTGG	1330
QY	1328	TCAACTATATTTCCCCAGGGTTCAATAGATTCACCTTTGGAAGAAAGTGTACTTAGTTA	1387
Db	1331	TCAACTATATTTCCCCAGGGTTCAATAGATTCACCTTTGGAAGAAAGTGTACTTAGTTA	1390
QY	1388	GGGTGATGAAAAGCACAGAAAGACTTACGCTGGCGCTCCCACTCCAATGATGACACTT	1447
Db	1391	GGGTGATGAAAAGCACAGAAAGACTTACGCTGGCGCTCCCACTCCAATGATGACACTT	1450
QY	1448	TGGATGAAGAGTAGAGCTTGTGCATTCAAAACATAGCAACTGAGCATATTCCAAGCCCTG	1507
Db	1451	TGGATGAAGAGTAGAGCTTGTGCATTCAAAACATAGCAACTGAGCATATTCCAAGCCCTG	1510

QY	1508	CCAAGAAAGTACCAAGACTTCCTCCCTGCCACAGACAGAGAACTTGAAAGCAGCTGTCAATTAGTA	1567
Db	1511	CCAAGAAAGTACCAAGACTTCCTCCCTGCCACAGACAGAGAACTTGAAAGCAGCTGTCAATTAGTA	1570
QY	1568	ATGGGAACATTAAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGCATGGGGT	1627
Db	1571	ATGGGAACATTAAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTGGGCATGGGGT	1630
QY	1628	GGGGTATGGGAACAGTTGG	1647
Db	1631	GGGGTATGGGAACAGTTGG	1650

RESULT 9				
US-10-236-417-197				
; Sequence 197, Application US/10236417				
; Publication No. US20040048256A1				
; GENERAL INFORMATION:				
; APPLICANT: Agee et al.				
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME				
; FILE REFERENCE: 21402-442C				
; CURRENT APPLICATION NUMBER: US/10/236,417				
; CURRENT FILING DATE: 2003-01-06				
; PRIOR APPLICATION NUMBER: US60/318,120				
; PRIOR FILING DATE: 2001-09-01				
; PRIOR APPLICATION NUMBER: US60/318,430				
; PRIOR FILING DATE: 2001-09-10				
; PRIOR APPLICATION NUMBER: US60/322,781				
; PRIOR FILING DATE: 2001-09-17				
; PRIOR APPLICATION NUMBER: US60/318,184				
; PRIOR FILING DATE: 2001-09-07				
; PRIOR APPLICATION NUMBER: US60/361,663				
; PRIOR FILING DATE: 2002-03-05				
; PRIOR APPLICATION NUMBER: US60/396,412				
; PRIOR FILING DATE: 2002-07-17				
; PRIOR APPLICATION NUMBER: US60/322,636				
; PRIOR FILING DATE: 2001-09-17				
; PRIOR APPLICATION NUMBER: US60/322,817				
; PRIOR FILING DATE: 2001-09-17				
; PRIOR APPLICATION NUMBER: US60/322,816				
; PRIOR FILING DATE: 2001-09-17				
; PRIOR APPLICATION NUMBER: US60/323,519				
; PRIOR FILING DATE: 2001-09-19				
; Remaining Prior Application data removed - See File Wrapper or PALM.				
; NUMBER OF SEQ ID NOS: 341				
; SOFTWARE: Custom				
; SEQ ID NO 197				
; LENGTH: 1650				
; TYPE: DNA				
; ORGANISM: Homo sapiens				
; FEATURE:				
; NAME/KEY: CDS				
; LOCATION: (22)..(1582)				
US-10-236-417-197				
Query Match				
Best Local Similarity 68.4%; Score 1370; DB 13; Length 1650;				
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;				
QY	134	GCTCTTCACCATGCTGATCACTTCTTTGAAATGCAGAGCTTGCGCCAAAAGATG	193	
Db	11	GCTCTTCACCATGCTGATCACTTCTTTGAAATGCAGAGCTTGCGCCAAAAGATG	70	
QY	194	TGGGAATGTTGCCCTTGAGATCTATTTTCTCTCAATATGTGATCAAGCAGAGTTGG	253	
Db	71	TTGGGATTTGTCCTTGAGATCTATTTTCTCTCAATATGTGATCAAGCAGAGTTGG	130	
QY	254	AAAATATGATGGTGTAGATGCTGGAAGATATCCATTGGCTTGGCCAGGCCAAGATGG	313	
Db	131	AAAATATGATGGTGTAGATGCTGGAAGATATCCATTGGCTTGGCCAGGCCAAGATGG	190	
QY	314	GCTTCTGACAGATAGAGAAGATTTAACTCTCTTGATGACTGTGTTCAAGAACTTTA	373	
Db				

Db 191 GCTTTCACACAGATAGAGAAGATATTAACTCTCTTTCAGTACTGTGGTTCAGAACTCTTA 250
QY 374 TGGAGAGAAATTAACCTTCTCTATGATTGTCATTGGCGGCTGGAAGTTGGAAACAGAGCAA 433
Db 251 TGGAGAGAAATTAACCTTCTCTATGATTGTCATTGGCGGCTGGAAGTTGGAAACAGAGCAA 310
QY 434 TCATCGACAAATCAAAAGCTGTGAAGACTAATTTGTATGTCAGCTGTTTGAAGAGCTGGGA 493
Db 311 TCATCGACAAATCAAAAGCTGTGAAGACTAATTTGTATGTCAGCTGTTTGAAGAGCTGGGA 370
QY 494 ATACAGATATAGAAGGAATCGACAACTAATGTCATGCTATGAGGACACAGCTGCTGCT 553
Db 371 ATACAGATATAGAAGGAATCGACAACTAATGTCATGCTATGAGGACACAGCTGCTGCT 430
QY 554 TCATGCTGTTAACTGGATTGAGTCCAGCTCTTGGAGT----- 591
Db 431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGAGTGAAGGATGCGGTATGCCGTGAGTTG 490
QY 592 ----- 591
Db 491 CAGGAGATATTGCTGTATATGCCACAGGAATGCTAGACCTACAGGTGAGTTGAGCAG 550
QY 592 -----GGGCTTCGTGGGACAC 607
Db 551 TAGCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACAC 610
QY 608 ATATGCAACATGCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 667
Db 611 ATATGCAACATGCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 670
QY 668 ATGGAATACTCTCCATACAGTGTCTAAGTGCATTTAGACCGCTGCTACTCTGTCTACT 727
Db 671 ATGGGAAACTCTCCATACAGTGTCTAAGTGCATTTAGACCGCTGCTACTCTGTCTACT 730
QY 728 GCAAAAAAGATCCATGCCCCCAGTGGCAGAAAAGAGGAAATGATAAGATTTTACCCTTGATG 787
Db 731 GCAAAAAAGATCCATGCCCCCAGTGGCAGAAAAGAGGAAATGATAAGATTTTACCCTTGATG 790
QY 788 ATTTGGCTTCATGATCTTTCACTCAACCATATTGTAACTGTTCAGAAATCTCTAGCTC 847
Db 791 ATTTGGCTTCATGATCTTTCACTCAACCATATTGTAACTGTTCAGAAATCTCTAGCTC 850
QY 848 GGATGTTGCTGAATGACTTCCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTG 907
Db 851 GGATGTTGCTGAATGACTTCCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTG 910
QY 908 GCCTGGAAGCCTTTGGGGAGTGTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA 967
Db 911 GCCTGGAAGCCTTTGGGGAGTGTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA 970
QY 968 AGGCATTATGAAGGCTAGCTCTGAACCTCTTCAGTCAGAAAACAAAGGCATCTTTACTTG 1027
Db 971 AGGCATTATGAAGGCTAGCTCTGAACCTCTTCAGTCAGAAAACAAAGGCATCTTTACTTG 1030
QY 1028 TATCAAAATCAAAATGAAATATGTACACATCTTCAGTATATGTTCCTTCATCTGTTC 1087
Db 1031 TATCAAAATCAAAATGAAATATGTACACATCTTCAGTATATGTTCCTTCATCTGTTC 1090
QY 1088 TAGCACAGTACTACCTCAGCAATTAGCAGGGAAGAAATTGAGTGTTCCTTATGTT 1147
Db 1091 TAGCACAGTACTACCTCAGCAATTAGCAGGGAAGAAATTGAGTGTTCCTTATGTT 1150
QY 1148 CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAGATGCTACACCGGGGTCTG 1207
Db 1151 CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAGATGCTACACCGGGGTCTG 1210
QY 1208 CTCTTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1267
Db 1211 CTCTTGATAAATTAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1270
QY 1268 GTGTGGCACCAGATGTCTTGGCTGAACCAATGAAGCTCAGAGAGGACACCCATCATTTGG 1327
Db 1271 GTGTGGCACCAGATGTCTTGGCTGAACCAATGAAGCTCAGAGAGGACACCCATCATTTGG 1330

QY 1328 TCAACTATATTCGCCAGGGTCAATAGATTCACTCTTGAAGGAACGTGTACTTACTTA 1387
Db 1331 TCAACTATATTCGCCAGGGTCAATAGATTCACTCTTGAAGGAACGTGTACTTACTTA 1390
QY 1388 GGGTGATGAAAAAGCAGAGAAGAACTTACGCTCGGCGCTCCCACTCCAAATGATGACTT 1447
Db 1391 GGGTGATGAAAAAGCAGAGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACTT 1450
QY 1448 TGGATGAAGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTCCAAGCCTG 1507
Db 1451 TGGATGAAGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTCCAAGCCTG 1510
QY 1508 CCAAGAAAGTACCAGACTCCCTGCCACAGCAGCAAGAACTGAAAGCAGCTGTCAATTA 1567
Db 1511 CCAAGAAAGTACCAGACTCCCTGCCACAGCAGCAAGAACTGAAAGCAGCTGTCAATTA 1570
QY 1568 ATGGGAACATTAAAGATCTGTGAGGTGCAAGACTTCAGGGTGGGTGGCATGGGGT 1627
Db 1571 ATGGGAACATTAAAGATCTGTGAGGTGCAAGACTTCAGGGTGGGTGGCATGGGGT 1630
QY 1628 GGGGTATGGGAACAGTTGG 1647
Db 1631 GGGGTATGGGAACAGTTGG 1650

RESULT 10
US-10-236-417-201
; Sequence 201, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 201
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1582)
US-10-236-417-201
Query Match 68.4%; Score 1370; DB 13; length 1650;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;
QY 134 GCTCTTCAACCATGCTGTGATCACTTCTTGAATGAGAAAGCTTGCTGGCCAAAGATG 193

Db 11 GCTCTTCAACCATGCGCTGATCACTTCTTGAAATGCAGAAAGCTTGCGCCAAAGATG 70
QY 194 TGGGAATGTGTCCTTGAGATCTATTTCTCTCAATATGTGATCAAGCAGATTGG 253
Db 71 TTGGGATGTGTCCTTGAGATCTATTTCTCTCAATATGTGATCAAGCAGATTGG 130
QY 254 AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTGCTGGCCAGGCCAAGATGG 313
Db 131 AAAAATATGATGGTGTAGATGCTGGGAAGTATACCATTTGGCTTGGCCAGGCCAAGATGG 190
QY 314 GCTTCTGCACAGATAGAGAAGATTTAACTCTCTTTCATGACTGTGTTGAGAACTTTA 373
Db 191 GCTTCTGCACAGATAGAGAAGATTTAACTCTCTTTCATGACTGTGTTGAGAACTTTA 250
QY 374 TGGAGAGAAATTAACCTTCTCTATGATTCGATTGGGCGGCTGGAAGTTGGAACAGACAA 433
Db 251 TGGAGAGAAATTAACCTTCTCTATGATTCGATTGGGCGGCTGGAAGTTGGAACAGACAA 310
QY 434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA 493
Db 311 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA 370
QY 494 ATACAGATATAGAAGGATCGACACAACTAATGCATGCTATGAGGCAAGCTGCTGCT 553
Db 371 ATACAGATATAGAAGGATCGACACAACTAATGCATGCTATGAGGCAAGCTGCTGCT 430
QY 554 TCAATGCTGTTAATGAGATTGAGTCCAGCTCTTGGGAT----- 591
Db 431 TCAATGCTGTTAATGAGATTGAGTCCAGCTCTTGGGATGGACGATGCCCTGTAGTTG 490
QY 592 ----- 591
Db 491 CAGGAGATATTGCTGATATGCCACAGGAATGCTAGACCTACAGGTGGAGTTGAGACAG 550
QY 592 -----GGGCTTCTGGGACAC 607
Db 551 TAGCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCTGTGGACAC 610
QY 608 ATATGCACATGCCCTATGATTTTACAAAGCTGATATGCTATCTGAATATCTATAGTAG 667
Db 611 ATATGCACATGCCCTATGATTTTACAAAGCTGATATGCTATCTGAATATCTATAGTAG 670
QY 668 ATGGAAACTCTCCATACAGTGTCTACCTGATGCATTAGACCGCTGCTACTCTGTACT 727
Db 671 ATGGAACTCTCCATACAGTGTCTACCTGATGCATTAGACCGCTGCTACTCTGTACT 730
QY 728 GCAAAAAGATCCATGCCAGTGGCAGAAAGAGGGAATGATTAAGATTTTACCTTGAATG 787
Db 731 GCAAAAAGATCCATGCCAGTGGCAGAAAGAGGGAATGATTAAGATTTTACCTTGAATG 790
QY 788 ATTTGGCTTCATGATCTTTCACTCCACATATTGTAACCTGGTTCAGAAATCTCTAGCTC 847
Db 791 ATTTGGCTTCATGATCTTTCACTCCACATATTGTAACCTGGTTCAGAAATCTCTAGCTC 850
QY 848 GGATGTTGCTGAATGACTTCCCTAATGACCAAGATAGAGATAAATAGTATCTATAGTG 907
Db 851 GGATGTTGCTGAATGACTTCCCTAATGACCAAGATAGAGATAAATAGTATCTATAGTG 910
QY 908 GCCTGGAAGCCTTTGGGATGTTAAATTAGAAACAACCTACTTTGATAGAGATGTGAGA 967
Db 911 GCCTGGAAGCCTTTGGGATGTTAAATTAGAAACAACCTACTTTGATAGAGATGTGAGA 970
QY 968 AGGCATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAAGGCATCTTTACTTG 1027
Db 971 AGGCATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAAGGCATCTTTACTTG 1030
QY 1028 TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTC 1087
Db 1031 TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTC 1090
QY 1088 TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAAATTGGAGTGTCTTATGCTT 1147

Db 1091 TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAAATTGAGTGTCTTATGCTT 1150
QY 1148 CTGGTTTGGCTGCACACTCTGTACTCTCTTAAAGTACACAAAGATGCTACACCGGGGTCTG 1207
Db 1151 CTGGTTTGGCTGCACACTCTGTACTCTCTTAAAGTACACAAAGATGCTACACCGGGGTCTG 1210
QY 1208 CTCTTGATTAATAACAGCAAGATTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1267
Db 1211 CTCTTGATTAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1270
QY 1268 GTGTGGCACCAGATGCTTCTGCTGAATAACATGAAAGCTCAGAGAGACACCCATCATTTGG 1327
Db 1271 GTGTGGCACCAGATGCTTCTGCTGAATAACATGAAAGCTCAGAGAGACACCCATCATTTGG 1330
QY 1328 TCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAAGTGTACTTAACTTA 1387
Db 1331 TCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAAGTGTACTTAACTTA 1390
QY 1388 GGGTGGATGAAAAGCACAGAAAGACTTACGCTCGGCGTCCCACTGCAATGATGACACTT 1447
Db 1391 GGGTGGATGAAAAGCACAGAAAGACTTACGCTCGGCGTCCCACTGCAATGATGACACTT 1450
QY 1448 TGGATGAAGAGTGAAGACTTGTGCACTTCAAAATAGCAACTGAGCAATTTCCAAAGCCCTG 1507
Db 1451 TGGATGAAGAGTGAAGACTTGTGCACTTCAAAATAGCAACTGAGCAATTTCCAAAGCCCTG 1510
QY 1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGACAGAACTGAAAGCAGCTGTCAATTAGTA 1567
Db 1511 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGACAGAACTGAAAGCAGCTGTCAATTAGTA 1570
QY 1568 ATGGGGAACATTAAAGTACTGTGTGAGGTGCAAGACTTCAAGGTGGGTGGCATGGGT 1627
Db 1571 ATGGGGAACATTAAAGTACTGTGTGAGGTGCAAGACTTCAAGGTGGGTGGCATGGGT 1630
QY 1628 GGGGTATGGGAACAGTTGG 1647
Db 1631 GGGGTATGGGAACAGTTGG 1650

RESULT 11
US-10-236-417-205
; Sequence 205, Application US/10236417
; Publication No. US20040048256a1
; GENERAL INFORMATION:
; APPLICANT: Agree et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 205

; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22) .. (1582)
US-10-236-417-205

Query Match 68.4%; Score 1370; DB 13; Length 1650;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

QY 134 GCTCTTTCACCATGCGCTGATCACTTCCTTGAATGCAGAGCTTGCGCCAAAAGATG 193
Db 11 GCTCTTTCACCATGCGCTGATCACTTCCTTGAATGCAGAGCTTGCGCCAAAAGATG 70
QY 194 TGGGAATTGTTGCCCTGAGATCTATTTTCTCTCAATATGTTGATCAAGCAGAGTTGG 253
Db 71 TTGGGATTGTTGCCCTGAGATCTATTTTCTCTCAATATGTTGATCAAGCAGAGTTGG 130
QY 254 AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCTGGCCAGGCCAAGATGG 313
Db 131 AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCTGGCCAGGCCAAGATGG 190
QY 314 GCTTCTGCACAGATAGAGAAGATATTAATCTCTCTTGACATGACTGTGTTCAAGATCTTA 373
Db 191 GCTTCTGCACAGATAGAGAAGATATTAATCTCTCTTGACATGACTGTGTTCAAGATCTTA 250
QY 374 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGCGGCTGGAAGTTGGAACAGAGACAA 433
Db 251 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGCGGCTGGAAGTTGGAACAGAGACAA 310
QY 434 TCATCGACAATCAAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGA 493
Db 311 TCATCGACAATCAAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGA 370
QY 494 ATACAGATATAGAAGGAATGCACAACATAATGCATGCTATGAGGACAGCTGCTGCT 553
Db 371 ATACAGATATAGAAGGAATGCACAACATAATGCATGCTATGAGGACAGCTGCTGCT 430
QY 554 TCAATGCTGTTAACTGATTTGATTCAGCTCTTGGGAT----- 591
Db 431 TCAATGCTGTTAACTGATTTGATTCAGCTCTTGGGATGGACGGTATGCCCTGTAAGTTG 490
QY 592 ----- 591
Db 491 CAGGAGATATTGCTGTATATGCCACAGGAATGCTAGACCTTACAGGTGGAGTGGAGCAG 550
QY 592 -----GGGCTTGTGGGAC 607
Db 551 TAGCTCTGCTAATTGGGCCAAATGCTCTTAAATTTTGAACGAGGGCTTGTGGGACAC 610
QY 608 ATATGCAACATGCGCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 667
Db 611 ATATGCAACATGCGCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 670
QY 668 ATGGAATACTCTCCATACAGTGTCACTCAGTGCATTAGACCGCTGCTACTCTGTCTACT 727
Db 671 ATGGAATACTCTCCATACAGTGTCACTCAGTGCATTAGACCGCTGCTACTCTGTCTACT 730
QY 728 GCAAAAAGATCCATGCCCCAGTGGCAGAAAGAGGAAATGATAAGATTTTACCCTGAATG 787
Db 731 GCAAAAAGATCCATGCCCCAGTGGCAGAAAGAGGAAATGATAAGATTTTACCCTGAATG 790
QY 788 ATTTGGCTTCATGATCTTTTCACTTCAACCATATTTGTAACCTGGTTCAGAAATCTTAGCTC 847
Db 791 ATTTGGCTTCATGATCTTTTCACTTCAACCATATTTGTAACCTGGTTCAGAAATCTTAGCTC 850
QY 848 GGAATGTTGCTGAATGACTTCCTTAATGACCAAGATAGAGATAAAAAATAGTATCTATAGTG 907
Db 851 GGAATGTTGCTGAATGACTTCCTTAATGACCAAGATAGAGATAAAAAATAGTATCTATAGTG 910
QY 908 GCCTGGAAGCCTTTGGGAGTGTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA 967

Db 911 GCCTGGAAGCCTTTGGGAGTGTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGA 970
QY 968 AGGCAATTTATGAAGGCTAGCTCTGAACCTCTTCAGTCAAGAAAACAAGGCATCTTACTTG 1027
Db 971 AGGCAATTTATGAAGGCTAGCTCTGAACCTCTTCAGTCAAGAAAACAAGGCATCTTACTTG 1030
QY 1028 TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGCTTCCCTTGCACTGTTC 1087
Db 1031 TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGCTTCCCTTGCACTGTTC 1090
QY 1088 TAGCAGACTACTCACTCAGCAATTAAGCAGGGAAGAAATGAGTGTTCCTTAATGCTT 1147
Db 1091 TAGCAGACTACTCACTCAGCAATTAAGCAGGGAAGAAATGAGTGTTCCTTAATGCTT 1150
QY 1148 CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAGATGCTACACCGGGTCTG 1207
Db 1151 CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAGATGCTACACCGGGTCTG 1210
QY 1208 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAGAACTG 1267
Db 1211 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAGAACTG 1270
QY 1268 GTGTGGCAACCAAGATGCTTTCGCTGAAAAACATGAAGCTCAAGAGGACACCCATCTTGG 1327
Db 1271 GTGTGGCAACCAAGATGCTTTCGCTGAAAAACATGAAGCTCAAGAGGACACCCATCTTGG 1330
QY 1328 TCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTACTTAGTTA 1387
Db 1331 TCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAACGTGTACTTAGTTA 1390
QY 1388 GGGTGGATGAAAAAGCACAGAGAATCTTACGCTCGGCGTCCCACTCCAATGATGACACTT 1447
Db 1391 GGGTGGATGAAAAAGCACAGAGAATCTTACGCTCGGCGTCCCACTCCAATGATGACACTT 1450
QY 1448 TGGATGAAGAGTAGACTGTGTGATTCAAAACATAGCACTGAGCATATTTCCAAGCCCTG 1507
Db 1451 TGGATGAAGAGTAGACTGTGTGATTCAAAACATAGCACTGAGCATATTTCCAAGCCCTG 1510
QY 1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAAGAACTGAAAGCAGCTGTCAATAGTA 1567
Db 1511 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAAGAACTGAAAGCAGCTGTCAATAGTA 1570
QY 1568 ATGGGAAACATTAAAGATCTCTGTGAGGTGCAAGACTTCAGGGTGGGTCATGGGCT 1627
Db 1571 ATGGGAAACATTAAAGATCTCTGTGAGGTGCAAGACTTCAGGGTGGGTCATGGGCT 1630
QY 1628 GGGGCTATGGGAACAGTTGG 1647
Db 1631 GGGGCTATGGGAACAGTTGG 1650

RESULT 12
US-10-307-817-359
; Sequence 359, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Aggee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307, 817
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: Curoseqlist version 0.1
; SEQ ID NO 359
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22) .. (1581)
US-10-307-817-359

Query Match		68.4%;	Score 1370;	DB 13;	Length 1650;						
Best Local Similarity		92.0%;	Pred. No. 0;								
Matches 1509;		Conservative 0;	Mismatches 5;	Indels 126;	Gaps 1;						
QY	134	GCTCTTCA	CCATG	CCCTG	ATCACTT	CTTGAATG	CAGAGC	CTTG	GGCCAA	AAGATG	193
Db	11	GCTCTTCA	CCATG	CCCTG	ATCACTT	CTTGAATG	CAGAGC	CTTG	GGCCAA	AAGATG	70
QY	194	TGGGAATG	TGGCC	TTGAG	ATCTAT	TTTCTC	TAATAT	TGTG	ATCAAG	CAGAGT	253
Db	71	TTGGGATG	TGCCCT	TGAGAT	CTATTT	CTCTCA	ATAATG	TTGAT	CAAGCA	GAGATT	130
QY	254	AAAAATAT	GATGT	GTAGAT	GTGAA	AGATAT	CCATTG	GCTT	GGCCAG	GC	313
Db	131	AAAAATAT	GATGT	GTAGAT	GTGAA	AGATAT	CCATTG	GCTT	GGCCAG	GC	190
QY	314	GCTTTCGA	CAGAT	AGAGA	GATAT	TAACTCT	CTTGC	ATGACT	GTG	CTTCA	373
Db	191	GCTTTCGA	CAGAT	AGAGA	GATAT	TAACTCT	CTTGC	ATGACT	GTG	CTTCA	250
QY	374	TGGAGAA	ATAAC	CTTTC	CTATG	ATG	TGGCG	CGCTG	GAAGT	TGGA	433
Db	251	TGGAGAA	ATAAC	CTTTC	CTATG	ATG	TGGCG	CGCTG	GAAGT	TGGA	310
QY	434	TCATGCA	CAATCA	AAAGCT	GTGA	AGACTA	TTTGAT	GACG	CTGTT	TGAAG	493
Db	311	TCATGCA	CAATCA	AAAGCT	GTGA	AGACTA	TTTGAT	GACG	CTGTT	TGAAG	370
QY	494	ATACAGAT	ATAGA	AGAA	TGCA	CACTA	TG	CATG	CTAT	GAGG	553
Db	371	ATACAGAT	ATAGA	AGAA	TGCA	CACTA	TG	CATG	CTAT	GAGG	430
QY	554	TCAATGCT	GTTAACT	GATG	AGTCC	AGCTCTT	GGAT	-----			591
Db	431	TCAATGCT	GTTAACT	GATG	AGTCC	AGCTCTT	GGAT	-----			490
QY	592	-----									591
Db	491	CAGAGAT	AT	TGCT	ATAT	GC	CA	CAGAA	TGCT	AGAC	550
QY	592	-----									607
Db	551	TAGCTCTG	CTAAT	TGGG	CCAA	TGCT	CTTT	TAAT	TTTGA	CGAGG	610
QY	608	ATATGCA	CAATG	CCCTAT	GAT	TTTT	TACA	AGCC	CTGAT	TGCT	667
Db	611	ATATGCA	CAATG	CCCTAT	GAT	TTTT	TACA	AGCC	CTGAT	TGCT	670
QY	668	ATGAAAA	ACTCC	ATACAG	TGCT	AC	AGT	GCAT	TAGAC	CGCTG	727
Db	671	ATGAAAA	ACTCC	ATACAG	TGCT	AC	AGT	GCAT	TAGAC	CGCTG	730
QY	728	GCAAAAA	GATCC	ATGCCC	AGTGC	AGAAAG	GGAA	TGAT	AAAG	ATTTA	787
Db	731	GCAAAAA	GATCC	ATGCCC	AGTGC	AGAAAG	GGAA	TGAT	AAAG	ATTTA	790
QY	788	ATTTTGG	CTTCA	TGAT	CTTT	CACTCA	CCAT	AT	TGTAA	CTGG	847
Db	791	ATTTTGG	CTTCA	TGAT	CTTT	CACTCA	CCAT	AT	TGTAA	CTGG	850
QY	848	GGATGTTG	CTGA	TGACT	TCCT	TAA	TAG	CA	GAAT	AGAT	907
Db	851	GGATGTTG	CTGA	TGACT	TCCT	TAA	TAG	CA	GAAT	AGAT	910
QY	908	GCCTGGA	AGCC	TTTGGG	ATGTT	AA	TTAG	AA	CA	CC	967
Db	911	GCCTGGA	AGCC	TTTGGG	ATGTT	AA	TTAG	AA	CA	CC	970
QY	968	AGGCAT	TTATGA	AGGCTAG	CT	CTGA	ACT	CTT	CA	GAA	1027
Db	971	AGGCAT	TTATGA	AGGCTAG	CT	CTGA	ACT	CTT	CA	GAA	1030
QY	1028	TATCA	AATCA	AAATG	GA	AATAT	GT	AC	AT	CT	1087

Db	1031	TATCAAAATCAAAATGGAATATATGTACACATCTTCAGTATATGGTTCCTTGCAATCTGTTCC	1090
QY	1088	TAGCA CAGTACTCA CCTCAGCAATTAGCAGGGAAGAATTTGGAGTGTCTTTCTTATGGTT	1147
Db	1091	TAGCA CAGTACTCA CCTCAGCAATTAGCAGGGAAGAATTTGGAGTGTCTTTCTTATGGTT	1150
QY	1148	CTGGTTTGGCTGCGCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTCTG	1207
Db	1151	CTGGTTTGGCTGCGCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTCTG	1210
QY	1208	CTCTTGATAAAAATACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG	1267
Db	1211	CTCTTGATAAAAATACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG	1270
QY	1268	GTGTGGCACCAGATGTCTTGGCTGAAAAATGAAGCTCAGAGAGACACCCATCATTTGG	1327
Db	1271	GTGTGGCACCAGATGTCTTGGCTGAAAAATGAAGCTCAGAGAGACACCCATCATTTGG	1330
QY	1328	TCAACTATATTTCCCCAGGGTTCAATAGATTCACTCTTTTGAAGGAACGTGTACTTAGTTA	1387
Db	1331	TCAACTATATTTCCCCAGGGTTCAATAGATTCACTCTTTTGAAGGAACGTGTACTTAGTTA	1390
QY	1388	GGGTGGATGAAAAGCACAGAAGACTTACGCTCGGCGTCCCACTCCAATGATGACACTT	1447
Db	1391	GGGTGGATGAAAAGCACAGAAGACTTACGCTCGGCGTCCCACTCCAATGATGACACTT	1450
QY	1448	TGGATGAAGAGTAGGACTTGTGCATTCAACATAGCAACTGAGCATATTTCCAAGCCCTG	1507
Db	1451	TGGATGAAGAGTAGGACTTGTGCATTCAACATAGCAACTGAGCATATTTCCAAGCCCTG	1510
QY	1508	CCAAGAAAAGTACCAAGACTCCCTGCGCACAGCAGCAGAACTGAAGCAGCTGTCAATTAGTA	1567
Db	1511	CCAAGAAAAGTACCAAGACTCCCTGCGCACAGCAGCAGAACTGAAGCAGCTGTCAATTAGTA	1570
QY	1568	ATGGGGAACATTAAAGTACTCTGTGAGGTGCAAGACTTCAGAGGTGGGTGGCATGGGGT	1627
Db	1571	ATGGGGAACATTAAAGTACTCTGTGAGGTGCAAGACTTCAGAGGTGGGTGGCATGGGGT	1630
QY	1628	GGGGGTATGGGAACAGTTGG 1647	
Db	1631	GGGGGTATGGGAACAGTTGG 1650	
RESULT 13			
US-10-307-817-371			
; Sequence 371, Application US/10307817			
; Publication No. US20040058338A1			
; GENERAL INFORMATION:			
; APPLICANT: Agee et al.			
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME			
; FILE REFERENCE: 21402-502C			
; CURRENT APPLICATION NUMBER: US/10/307, 817			
; CURRENT FILING DATE: 2002-12-02			
; NUMBER OF SEQ ID NOS: 682			
; SOFTWARE: CuroSeqIst version 0.1			
; SEQ ID NO 371			
; LENGTH: 1650			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (22) .. (1581)			
US-10-307-817-371			
Query Match 68.4%; Score 1370; DB 13; Length 1650;			
Best Local Similarity 92.0%; Pred. No. 0;			
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1			
QY	134	GCTCTTTCA CCAATG CCGTGAATCACTTCCTTTGAATGCAGAGCTTGCTGGCCAAAGATG	193
Db	11	GCTCTTTCA CCAATG CCGTGAATCACTTCCTTTGAATGCAGAGCTTGCTGGCCAAAGATG	70

QY	194	TGGGAATTTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGATTGG	253
Db	71	TTGGGATTGTGGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGATTGG	130
QY	254	AAAAATATGATGGTGTAGATGCTGGAAAGTATACCAATGGCTTGGCCAGGCCAAGATGG	313
Db	131	AAAAATATGATGGTGTAGATGCTGGGAAGTATACCAATGGCTTGGCCAGGCCAAGATGG	190
QY	314	GCTTCTGCACAGATAGAGAAGATATTTACTCTCTTTGCATGACTGTGGTTCAGAATCTTA	373
Db	191	GCTTCTGCACAGATAGAGAAGATATTTACTCTCTTTGCATGACTGTGGTTCAGAATCTTA	250
QY	374	TGGAGAGAAATAACCTTTCTCTATGATTTGCATTGGGCGGCTGGAAGTTGGAACAGACAA	433
Db	251	TGGAGAGAAATAACCTTTCTCTATGATTTGCATTGGGCGGCTGGAAGTTGGAACAGACAA	310
QY	434	TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA	493
Db	311	TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA	370
QY	494	ATACAGATATAGAAGAAATCCACACAACTAATGCATGCTATGAGGACACAGCTGCTCT	553
Db	371	ATACAGATATAGAAGAAATCCACACAACTAATGCATGCTATGAGGACACAGCTGCTCT	430
QY	554	TCAATGCTGTTAACTGGAATGAGTCCAGCTCTTGGAT-----	591
Db	431	TCAATGCTGTTAACTGGAATGAGTCCAGCTCTTGGAGTGAACGGTATGCCCTGTAGTTG	490
QY	592	-----	591
Db	491	CAGGAGATATTGCTGTATATGCCACAGAAATGCTAGACTACAGTGAAGTTGAGCAG	550
QY	592	-----GGGCTTCGTGGGACAC	607
Db	551	TAGCTCTGCTAATTGGGCCAAATGCTCTCTTAATTTTGAACGAGGGCTTCGTGGGACAC	610
QY	608	ATATGCACACATGCCCTATGATTTTTCACAAGCCTGATATGCTATCTGAATATCCTATAGTAG	667
Db	611	ATATGCACACATGCCCTATGATTTTTCACAAGCCTGATATGCTATCTGAATATCCTATAGTAG	670
QY	668	ATGAAAACTCTCCATPACAGTGTCACTCACTCAGTGCATTAGAACCGCTGCTACTCTGTCTACT	727
Db	671	ATGGAACACTCTCCATPACAGTGTCACTCACTCAGTGCATTAGAACCGCTGCTACTCTGTCTACT	730
QY	728	GCAAAAAGATCCATGCCCACTGGCAGAAAAGGGGAATGATAAAGATTTTACCCTTGAATG	787
Db	731	GCAAAAAGATCCATGCCCACTGGCAGAAAAGGGGAATGATAAAGATTTTACCCTTGAATG	790
QY	788	ATTTTGGCTTCATGATCTTTTCACTCACCATATTGTAACTGGTTCAGAAATCTCTAGCTC	847
Db	791	ATTTTGGCTTCATGATCTTTTCACTCACCATATTGTAACTGGTTCAGAAATCTCTAGCTC	850
QY	848	GGATGTTGCTGAATGACTTCTTAATGACCAGAATAGATAAAAATAGTATCTATAGTG	907
Db	851	GGATGTTGCTGAATGACTTCTTAATGACCAGAATAGATAAAAATAGTATCTATAGTG	910
QY	908	GCCTGGAAGCCTTTGGGATGTTAAATTGAAGACACCTTCTTGTATAGAGATGTGAGAA	967
Db	911	GCCTGGAAGCCTTTGGGATGTTAAATTGAAGACACCTTCTTGTATAGAGATGTGAGAA	970
QY	968	AGGCATTATGAAGGCTAGCTCTGAACCTTTCAGTTCAGAAAACAAAGGCATCTTTACTTG	1027
Db	971	AGGCATTATGAAGGCTAGCTCTGAACCTTTCAGTTCAGAAAACAAAGGCATCTTTACTTG	1030
QY	1028	TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATAGGTTCCCTTGACATCTGTC	1087
Db	1031	TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATAGGTTCCCTTGACATCTGTC	1090
QY	1088	TAGCACAGTACTCACCTTCAGCAATTAGCAGGGAAGAGAAATTGAGTGTCTTCTATGCTT	1147
Db	1091	TAGCACAGTACTCACCTTCAGCAATTAGCAGGGAAGAGAAATTGAGTGTCTTCTATGCTT	1150
QY	1148	CTGGTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGCTCTG	1207

```

Db      1151  CTGGTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTCTG 1210
QY      1208  CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1267
Db      1211  CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1270
QY      1268  GTGTGGCACCAGATGTCCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1327
Db      1271  GTGTGGCACCAGATGTCCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1330
QY      1328  TCAACTATATTTCCCGAGGGTTCAATAGATTCACCTTTTGAAGAACTGTACTTAGTTA 1387
Db      1331  TCAACTATATTTCCCGAGGGTTCAATAGATTCACCTTTTGAAGAACTGTACTTAGTTA 1390
QY      1388  GGGTGATGAAAAAGCACAGAAAGAACTTACGCTCGCGCTCCCACTCCAAATGATGACACTT 1447
Db      1391  GGGTGATGAAAAAGCACAGAAAGAACTTACGCTCGCGCTCCCACTCCAAATGATGACACTT 1450
QY      1448  TGGATGAGGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCGAAGCCCTG 1507
Db      1451  TGGATGAGGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCGAAGCCCTG 1510
QY      1508  CCAAGAAAGTACCAGAAGCTCCCTGCCACAGACAGCAAGAACTGAAGCAGCTGTATTAGTA 1567
Db      1511  CCAAGAAAGTACCAGAAGCTCCCTGCCACAGACAGCAAGAACTGAAGCAGCTGTATTAGTA 1570
QY      1568  ATGGGGCAATTAGATACTCTGTGTGAGGTGCAAGAATTCAAGGTGGGGTGGGCATGGGGT 1627
Db      1571  ATGGGGCAATTAGATACTCTGTGTGAGGTGCAAGAATTCAAGGTGGGGTGGGCATGGGGT 1630
QY      1628  GGGGGTATGGGAACAGTTGG 1647
Db      1631  GGGGGTATGGGAACAGTTGG 1650

RESULT 14
US-10-307-817-445
; Sequence 445, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Agsee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307,817
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 445
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-307-817-445

Query Match      68.4%; Score 1370; DB 13; Length 1650;
Best local Similarity 92.0%; Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

QY      134  GCTCTTTCACCATGCGCTGGATCACTTCCTTTGAATGACAGAAAGCTTGCGCCAAAAGATG 193
Db      11  GCTCTTTCACCATGCGCTGGATCACTTCCTTTGAATGACAGAAAGCTTGCGCCAAAAGATG 70
QY      194  TGGGAATGTGTGCGCTTGAGATCTATTTTCTTCAATATGTGTGATCAAGAGATTGG 253
Db      71  TTGGGATGTGTGCGCTTGAGATCTATTTTCTTCAATATGTGTGATCAAGAGATTGG 130
QY      254  AAAAATATGATGTTAGTGTAGATGCTGGAAGTATACCAATTGGCTTGCGCCAGGCCAAGATGG 313
Db      131  AAAAATATGATGTTAGTGTAGATGCTGGAAGTATACCAATTGGCTTGCGCCAGGCCAAGATGG 190
QY      314  GCTTTCGACAGATAGAGAAAGATATTAATCTCTTTTCATGACTGTGGTTCAGAATCTTA 373
Db      191  GCTTTCGACAGATAGAGAAAGATATTAATCTCTTTTCATGACTGTGGTTCAGAATCTTA 250

```


QY 374 TGGAGAGAAATAACCTTTCCTATGATTCGATTCGGCGGCTGGAAGTTGGAACAGACAA 433
| | | | |
Db 251 TGGAGAGAAATAACCTTTCCTATGATTCGATTCGGCGGCTGGAAGTTGGAACAGACAA 310
QY 434 TCATCGACAATCAAAGTCTGTGAAGACTAATTGATGACGTGTTTGAAGACTCTGGGA 493
| | | | |
Db 311 TCATCGACAATCAAAGTCTGTGAAGACTAATTGATGACGTGTTTGAAGACTCTGGGA 370
QY 494 ATACAGATATAGAAGAAATCGACACAACCTAATGCTATGAGGCGACAGCTGCTGCT 553
| | | | |
Db 371 ATACAGATATAGAAGAAATCGACACAACCTAATGCTATGAGGCGACAGCTGCTGCT 430
QY 554 TCAATGCTGTAACTGGAATTGAGTCCAGCTCTTGGGAT----- 591
| | | | |
Db 431 TCAATGCTGTAACTGGAATTGAGTCCAGCTCTTGGGATGAGCGGTATGCCCTGTAGTTG 490
QY 592 ----- 591
Db 491 CAGAGATATTGCTGTATATGCCACAGAAATGCTAGACCTACAGGTGGAGTTGAGCAG 550
QY 592 -----GGGCTTCGTGGACAC 607
| | | | |
Db 551 TAGCTCTGCTAATTGGGCCCAATGCTCTTAAATTTTGAACGAGGCTTCGTGGACAC 610
QY 608 ATATGCAACATGCGCTATGATTTTTCAGAGCCTGATATGCTATCTGAATATCCTATAGTAG 667
| | | | |
Db 611 ATATGCAACATGCGCTATGATTTTTCAGAGCCTGATATGCTATCTGAATATCCTATAGTAG 670
QY 668 ATGGAATACTCTCCATACAGTGCTACTCAGTGATAGACCGCTGCTACTGTGCTACT 727
| | | | |
Db 671 ATGGGAAACTCTCCATACAGTGCTACTCAGTGATAGACCGCTGCTACTGTGCTACT 730
QY 728 GCAAAAAGATCCATGCCAGTGGCAGAAAAGGAAATGATAAAGATTTTACCCTTGAATG 787
| | | | |
Db 731 GCAAAAAGATCCATGCCAGTGGCAGAAAAGGAAATGATAAAGATTTTACCCTTGAATG 790
QY 788 ATTTGGCTTCATGATCTTTCACCTCACCATATTTGTAACTGGTTCAGAAATCTCTAGCTC 847
| | | | |
Db 791 ATTTGGCTTCATGATCTTTCACCTCACCATATTTGTAACTGGTTCAGAAATCTCTAGCTC 850
QY 848 GGATGTGCTGAATGACTCTCTTAATGACCAAGATAGAGATAAATATGATCTATAGTG 907
| | | | |
Db 851 GGATGTGCTGAATGACTCTCTTAATGACCAAGATAGAGATAAATATGATCTATAGTG 910
QY 908 GCCTGGAAGCCTTTGGGAGTGTAAATTAGAGACACCTACTTTGATAGAGATGTGAGA 967
| | | | |
Db 911 GCCTGGAAGCCTTTGGGAGTGTAAATTAGAGACACCTACTTTGATAGAGATGTGAGA 970
QY 968 AGGCATTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAGGCATCTTTACTTG 1027
| | | | |
Db 971 AGGCATTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAGGCATCTTTACTTG 1030
QY 1028 TATCAATCAAAATGGAATATGTACACATCTTCAGTATATGCTTCCCTTGCAATCTGTT 1087
| | | | |
Db 1031 TATCAATCAAAATGGAATATGTACACATCTTCAGTATATGCTTCCCTTGCAATCTGTT 1090
QY 1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAAATGGAGTGTCTTATAGTT 1147
| | | | |
Db 1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAAATGGAGTGTCTTATAGTT 1150
QY 1148 CTGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAGATGCTACACCGGGGCTCTG 1207
| | | | |
Db 1151 CTGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAGATGCTACACCGGGGCTCTG 1210
QY 1208 CTCTTGATAAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1267
| | | | |
Db 1211 CTCTTGATAAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTG 1270
QY 1268 GTGTGGCAACCAGATGTCTTGCCTGAAAAACATGAAGCTCAGAGAGGACACCCCATATTGG 1327
| | | | |
Db 1271 GTGTGGCAACCAGATGTCTTGCCTGAAAAACATGAAGCTCAGAGAGGACACCCCATATTGG 1330

QY 1328 TCAACTATATTCGCCAGGGTTCAATAGATTCACTCTTTGAAGGAACTGTACTTAGTTA 1387
| | | | |
Db 1331 TCAACTATATTCGCCAGGGTTCAATAGATTCACTCTTTGAAGGAACTGTACTTAGTTA 1390
QY 1388 GGGTGGATGAAAAAGCACAGAGAACTTACGCTCGCGCTCCCACTCCAATGATGACACTT 1447
| | | | |
Db 1391 GGGTGGATGAAAAAGCACAGAGAACTTACGCTCGCGCTCCCACTCCAATGATGACACTT 1450
QY 1448 TGGATGAAGAGTAGGACTGTGTGCAATTCAAACATAGCAACTGAGCATATTCCAGCCCTG 1507
| | | | |
Db 1451 TGGATGAAGAGTAGGACTGTGTGCAATTCAAACATAGCAACTGAGCATATTCCAGCCCTG 1510
QY 1508 CCAAGAAAGTACCAAGACTCCCTGCGCACAGCAGAGAACTGAAAGCACTGTCAATAGTA 1567
| | | | |
Db 1511 CCAAGAAAGTACCAAGACTCCCTGCGCACAGCAGAGAACTGAAAGCACTGTCAATAGTA 1570
QY 1568 ATGGGGAACATTAAATACTCTGTGTGAGGTGCAAGACTTCAGGGTGGGTGGCATGGGCT 1627
| | | | |
Db 1571 ATGGGGAACATTAAATACTCTGTGTGAGGTGCAAGACTTCAGGGTGGGTGGCATGGGCT 1630
QY 1628 GGGGCTATGGGAACGTTGG 1647
| | | | |
Db 1631 GGGGCTATGGGAACGTTGG 1650

RESULT 15
US-10-354-358-55
; Sequence 55, Application US/10354358
; Publication No. US20030157082A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc
; APPLICANT: Hunter, John Joseph
; APPLICANT: Macbeth, Kyle J.
; APPLICANT: Tsai, Fong-Ying
; APPLICANT: Lesoon, Andrea
; APPLICANT: Lightcap, Eric S.
; APPLICANT: Williamson, Mark
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING
; TITLE OF INVENTION: CANCER USING 140, 1470, 1686, 2089, 2427, 3702, 5891, 6428,
; TITLE OF INVENTION: 7181, 7660, 25641, 69583, 49863, 8897, 1682, 17667, 9235,
; TITLE OF INVENTION: 3703, 14171, 10359, 1660, 1450, 18894, 2088, 32427, 2160,
; TITLE OF INVENTION: 9252, 9389, 1642, 85269, 10297, 1584, 9525, 14124, 4469,
; TITLE OF INVENTION: 8990, 2100, 9288, 64698, 10480, 20893, 33230, 1586, 9943,
; TITLE OF INVENTION: 16334, 68862, 9011, 14031, 6178, 21225, 1420, 32236, 2099,
; TITLE OF INVENTION: 2150, 26583, 2784, 8941, 9811, 27444, 50566 OR 66428 MOLECULES
; FILE REFERENCE: MP102-020P1R0M0NIM
; CURRENT APPLICATION NUMBER: US/10/354,358
; CURRENT FILING DATE: 2003-01-30
; PRIOR APPLICATION NUMBER: US 60/353,600
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/364,517
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/371,075
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: US 60/371,507
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: US 60/372,984
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: US 60/374,194
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/382,995
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/385,023
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/388,853
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/389,395
; PRIOR FILING DATE: 2002-06-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 55

```

; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22) ... (1584)
; US-10-354-358-55

```

Query Match	68.4%	Score 1370;	DB 15;	Length 1650;
Best Local Similarity	92.0%	Pred. No. 0;		
Matches 1509; Conservative	0;	Mismatches 5;	Indels 126;	Gaps 1;

QY	134	GCTCTTTCACCATGCGCTGGATCATCTTCCTTGAATGCAGAAAGCTTGCTGGCCAAAAGATG	193
Db	11	GCTCTTTCACCATGCGCTGGATCATCTTCCTTGAATGCAGAAAGCTTGCTGGCCAAAAGATG	70
QY	194	TGGGAATGTGTGCCCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGACAGATTGG	253
Db	71	TTGGGATGTGTGCCCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGACAGATTGG	130
QY	254	AAAAATATGATGAGTGTAGATGCTGGAAGTATACCATTTGGCTTGGCCAGGCCAAGATGG	313
Db	131	AAAAATATGATGAGTGTAGATGCTGGAAGTATACCATTTGGCTTGGCCAGGCCAAGATGG	190
QY	314	GCTTCTGCACAGATAGAGAATATTACTCTCTTTCATGACTGTGGTTCAGAATCTTA	373
Db	191	GCTTCTGCACAGATAGAGAATATTACTCTCTTTCATGACTGTGGTTCAGAATCTTA	250
QY	374	TGGAGAGAAATAACCTTTCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGACAA	433
Db	251	TGGAGAGAAATAACCTTTCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGACAA	310
QY	434	TCATCGACAAATCAAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTGAAGACTGGGA	493
Db	311	TCATCGACAAATCAAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTGAAGACTGGGA	370
QY	494	ATACAGATATAGAAGGAATCGACACAACATAATGCATGCTATGAGGACACAGCTGCTCT	553
Db	371	ATACAGATATAGAAGGAATCGACACAACATAATGCATGCTATGAGGACACAGCTGCTCT	430
QY	554	TCAATGCTGTTAACTGATGATTGAGTCCAGCTCTTGGGAT-----	591
Db	431	TCAATGCTGTTAACTGATTGAGTCCAGCTCTTGGGATGGAACGGTATGCCCTGGTAGTTG	490
QY	592	-----	591
Db	491	CAGAGATATGCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGCAG	550
QY	592	-----GGGCTTCGTGGGACAC	607
Db	551	TAGCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACAC	610
QY	608	ATATGCAACATGCGCTATGATTTTTCACAAGCCTGATATGCTATCTGAATATCTTAAGTAG	667
Db	611	ATATGCAACATGCGCTATGATTTTTCACAAGCCTGATATGCTATCTGAATATCTTAAGTAG	670
QY	668	ATGGAATACTCTCCATACAGTGTCTACTCAGTGCAATTAGACCGCTGCTACTCTGTCTACT	727
Db	671	ATGGAATACTCTCCATACAGTGTCTACTCAGTGCAATTAGACCGCTGCTACTCTGTCTACT	730
QY	728	GCAAAAAAGATCCATGCCCCAGTGGCAGAAAAGAGGAAATGATAAAGATTTTAACTTGAATG	787
Db	731	GCAAAAAAGATCCATGCCCCAGTGGCAGAAAAGAGGAAATGATAAAGATTTTAACTTGAATG	790
QY	788	ATTTTGGCTTCATGATCTTTCACTCAACCATATTGTAAACTGTTCAGAAATCTTAGCTC	847
Db	791	ATTTTGGCTTCATGATCTTTCACTCAACCATATTGTAAACTGTTCAGAAATCTTAGCTC	850
QY	848	GGAATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAAAAATAGTATCTATAGTG	907
Db	851	GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAAAAATAGTATCTATAGTG	910
QY	908	GCCCTGGAAGCCTTTGGGGATGTTAAATTGAAGACACCTACTTTGATAGAGATGTGGAGA	967

Db	911	GCCTGGAAGCCTTGGGGATGTTAAATTAGAGACACCTACTTTGATAGAGATGTGAGA	970
QY	968	AGGCATTTATGAAGGCTAGCTCTGAAGCTCTTCAATCAGAAAACAAAGGCATCTTTACTTG	1027
Db	971	AGGCATTTATGAAGGCTAGCTCTGAAGCTCTTCAATCAGAAAACAAAGGCATCTTTACTTG	1030
QY	1028	TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTTC	1087
Db	1031	TATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTTC	1090
QY	1088	TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAAATTGGAGTGTTCCTTAATGCGTT	1147
Db	1091	TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAAATTGGAGTGTTCCTTAATGCGTT	1150
QY	1148	CTGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAGATGCTACACCGGGCTCTG	1207
Db	1151	CTGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACAAGATGCTACACCGGGCTCTG	1210
QY	1208	CTCTTGATAAATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAGAACTG	1267
Db	1211	CTCTTGATAAATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAGAACTG	1270
QY	1268	GTGTGGCACCAAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGACACCCCATCATTTGG	1327
Db	1271	GTGTGGCACCAAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGACACCCCATCATTTGG	1330
QY	1328	TCAACTATATTCCTCCAGGGTTCATATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA	1387
Db	1331	TCAACTATATTCCTCCAGGGTTCATATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA	1390
QY	1388	GGGTGATGAAAAACAAGAAAGCTTACGCTCGGCGTCCCACTCCAAATGATGACACTT	1447
Db	1391	GGGTGATGAAAAACAAGAAAGCTTACGCTCGGCGTCCCACTCCAAATGATGACACTT	1450
QY	1448	TGGATGAAGGAGTAGGACTTGTGCATTCAAAATAGCAACTGAGCATATTTCAAGCCCTG	1507
Db	1451	TGGATGAAGGAGTAGGACTTGTGCATTCAAAATAGCAACTGAGCATATTTCAAGCCCTG	1510
QY	1508	CCAAGAAAGTACCAAGACTCCCTGCAACAGCAGCAGAACTGGAAGCAGCTGTCAATTA	1567
Db	1511	CCAAGAAAGTACCAAGACTCCCTGCAACAGCAGCAGAACTGGAAGCAGCTGTCAATTA	1570
QY	1568	ATGGGAACATTAAGATACTCTGTGAGGTGACAGACTTCAGGGTGGGGTGGGCAATGGGGT	1627
Db	1571	ATGGGAACATTAAGATACTCTGTGAGGTGACAGACTTCAGGGTGGGGTGGGCAATGGGGT	1630
QY	1628	GGGGGTATGGGAACAGTTGG 1647	
Db	1631	GGGGGTATGGGAACAGTTGG 1650	

Search completed: June 24, 2004, 13:16:16
Job time : 860 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 24, 2004, 13:20:33 ; Search time 883 Seconds
(without alignments)
152.827 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPGLPLNAEACWPKDVGIV.....PRLPATTAEPBAVISNGEH 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2511	100.0	478	13 US-10-193-295-2	Sequence 2, Appli
2	2511	100.0	478	15 US-10-622-516-2	Sequence 2, Appli
3	2480	98.8	520	12 US-10-236-417-182	Sequence 182, App
4	2480	98.8	520	12 US-10-236-417-184	Sequence 184, App
5	2480	98.8	520	12 US-10-236-417-186	Sequence 186, App
6	2480	98.8	520	12 US-10-236-417-190	Sequence 190, App
7	2480	98.8	520	12 US-10-236-417-194	Sequence 194, App
8	2480	98.8	520	12 US-10-236-417-198	Sequence 198, App
9	2480	98.8	520	12 US-10-236-417-202	Sequence 202, App
10	2480	98.8	520	12 US-10-236-417-204	Sequence 204, App
11	2480	98.8	520	12 US-10-236-417-206	Sequence 206, App
12	2480	98.8	520	12 US-10-236-417-208	Sequence 208, App
13	2480	98.8	520	12 US-10-307-817-360	Sequence 360, App
14	2480	98.8	520	12 US-10-307-817-370	Sequence 370, App
15	2480	98.8	520	12 US-10-307-817-372	Sequence 372, App

16	2480	98.8	520	12 US-10-307-817-374	Sequence 374, App
17	2480	98.8	520	12 US-10-307-817-467	Sequence 467, App
18	2480	98.8	520	12 US-10-307-817-468	Sequence 468, App
19	2480	98.8	520	12 US-10-041-018-292	Sequence 292, App
20	2480	98.8	520	13 US-10-193-295-4	Sequence 4, Appli
21	2480	98.8	520	14 US-10-354-358-56	Sequence 56, Appl
22	2480	98.8	520	15 US-10-622-516-4	Sequence 4, Appli
23	2480	98.8	529	12 US-10-236-417-192	Sequence 192, App
24	2480	98.8	529	12 US-10-307-817-364	Sequence 364, App
25	2475	98.6	519	12 US-10-236-417-188	Sequence 188, App
26	2475	98.6	519	12 US-10-307-817-362	Sequence 362, App
27	2475	98.6	524	12 US-10-236-417-196	Sequence 196, App
28	2475	98.6	524	12 US-10-236-417-212	Sequence 212, App
29	2475	98.6	524	12 US-10-307-817-366	Sequence 366, App
30	2475	98.6	524	12 US-10-307-817-376	Sequence 376, App
31	2475	98.6	525	12 US-10-236-417-200	Sequence 210, App
32	2475	98.6	525	12 US-10-307-817-210	Sequence 210, App
33	2475	98.6	525	12 US-10-307-817-358	Sequence 358, App
34	2475	98.6	525	12 US-10-307-817-368	Sequence 368, App
35	2439	97.1	518	13 US-10-193-295-5	Sequence 5, Appli
36	2439	97.1	518	15 US-10-622-516-5	Sequence 5, Appli
37	2362	94.1	520	12 US-10-205-331-14	Sequence 14, Appl
38	2362	94.1	520	12 US-10-041-018-313	Sequence 313, App
39	2352	93.7	520	12 US-10-041-018-314	Sequence 314, App
40	2097	83.5	522	12 US-10-041-018-315	Sequence 315, App
41	1513	60.3	508	12 US-10-041-018-296	Sequence 296, App
42	1513	60.3	508	14 US-10-205-823-174	Sequence 174, App
43	1513	60.3	508	14 US-10-177-293-208	Sequence 208, App
44	1494	59.5	508	12 US-10-041-018-316	Sequence 316, App
45	1493	59.5	480	12 US-10-041-018-284	Sequence 284, App

ALIGNMENTS

RESULT 1
US-10-193-295-2
; Sequence 2, Application US/10193295
; Publication No. US20020173018A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195DIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
US-10-193-295-2

Query Match	100.0%;	Score 2511;	DB 13;	Length 478;
Best Local Similarity	100.0%;	Pred. No. 5.2e-241;		
Matches 478;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKXVDAGKYTIGLGOAKMGFCT	60	
Db	1	MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKXVDAGKYTIGLGOAKMGFCT	60	
QY	61	DREDINSLCMTVVQNTLMERNNISYDCIGRLEVGTEITIDSKSVKTNLMQLFEESGNTDI	120	
Db	61	DREDINSLCMTVVQNTLMERNNISYDCIGRLEVGTEITIDSKSVKTNLMQLFEESGNTDI	120	
QY	121	EGIDTTNACYGTAAVFNNAVNTIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDKLSI	180	
Db	121	EGIDTTNACYGTAAVFNNAVNTIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDKLSI	180	

Fri Jun 25 07:33:53 2004

us-10-622-516-2.rappb

Page 2

QY	181	QCYL\$ALDRCY\$VYCKKIHAQ\$OKEGNDKDFTLNDEGFMI FHS\$PYCKLVQ\$SLARMLND	240
Db	181	QCYL\$ALDRCY\$VYCKKIHAQ\$OKEGNDKDFTLNDEGFMI FHS\$PYCKLVQ\$SLARMLND	240
QY	241	FLNDQNRDKNSIY\$GLEAF\$GVDYLEDITYFDRDVEKA\$FMKASSELFSQ\$KTKASLLVSNQNG	300
Db	241	FLNDQNRDKNSIY\$GLEAF\$GVDYLEDITYFDRDVEKA\$FMKASSELFSQ\$KTKASLLVSNQNG	300
QY	301	NMYTSSVY\$GLASVLAQ\$Y\$PQOLAGKRIGVFSY\$SGHAATLY\$SLKVTQDAPPG\$ALDKIT	360
Db	301	NMYTSSVY\$GLASVLAQ\$Y\$PQOLAGKRIGVFSY\$SGHAATLY\$SLKVTQDAPPG\$ALDKIT	360
QY	361	ASL\$CDL\$SRLD\$SRRTGVA\$PDVFA\$ENMKL\$REDTHL\$VNIYI\$PQ\$SIDSL\$FEGTWYLV\$RVDEKH	420
Db	361	ASL\$CDL\$SRLD\$SRRTGVA\$PDVFA\$ENMKL\$REDTHL\$VNIYI\$PQ\$SIDSL\$FEGTWYLV\$RVDEKH	420
QY	421	RRTYARRP\$TPND\$DTL\$DEGVGLV\$HSNIATEHI\$P\$SPAKVY\$PRL\$PATAAE\$PEAAV\$ISNGEH	478
Db	421	RRTYARRP\$TPND\$DTL\$DEGVGLV\$HSNIATEHI\$P\$SPAKVY\$PRL\$PATAAE\$PEAAV\$ISNGEH	478

```

RESULT 2
US-10-622-516-2
; Sequence 2, Application US/10622516
; Publication No. US20040018545A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195DIV2
; CURRENT APPLICATION NUMBER: US/10/622,516
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 10/193,295
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
US-10-622-516-2

```

	Query Match	100.0%;	Score 2511;	DB 15;	length 478;	
	Best Local Similarity	100.0%;	Pred. No. 5.2e-241;			
	Matches 478;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	MPGSLPINAACWPKDVGIVALEIYFSPQYVDQAELEKXVDGADGAKYTTIGLGOAKMGFCT	60			
Db	1	MPGSLPINAACWPKDVGIVALEIYFSPQYVDQAELEKXVDGADGAKYTTIGLGOAKMGFCT	60			
QY	61	DREDINSLCMTVVQNLMERNNLSYDCIGRLVEGTETIIDKSKSVKTNLMQLFEESGNTDI	120			
Db	61	DREDINSLCMTVVQNLMERNNLSYDCIGRLVEGTETIIDKSKSVKTNLMQLFEESGNTDI	120			
QY	121	EGIDDTNACYGGTAAVENAVNNWIESSWDGLRGTHMOHAYDFFKPPDMLSEXPVVDGKLSI	180			
Db	121	EGIDDTNACYGGTAAVENAVNNWIESSWDGLRGTHMOHAYDFFKPPDMLSEXPVVDGKLSI	180			
QY	181	QCYLSALDRCYSVYCKKIIHAWQWKEGNDKDFTLNDFGFMI FHS PYCKLVQKSLARMLND	240			
Db	181	QCYLSALDRCYSVYCKKIIHAWQWKEGNDKDFTLNDFGFMI FHS PYCKLVQKSLARMLND	240			
QY	241	FLNDQNRDKNSIYSGLEAFGVDVLEDTYFDRDVEKAFMKASSELFSQKTKASILLVSNONG	300			
Db	241	FLNDQNRDKNSIYSGLEAFGVDVLEDTYFDRDVEKAFMKASSELFSQKTKASILLVSNONG	300			
QY	301	NMYTSSVYGSILASVLAQYSPQOLAGKRIGVFSYSGLAATLYSLKVTQDATPQSALDKIT	360			
Db	301	NMYTSSVYGSILASVLAQYSPQOLAGKRIGVFSYSGLAATLYSLKVTQDATPQSALDKIT	360			

[illegible]

```

RESULT 3
; US-10-236-417-182
; Sequence 182, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agsee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 182
; LENGTH: 520
; TYPE: PRI
; ORGANISM: Homo sapiens
;
US-10-236-417-182

```

Query Match	98.8%;	Score 2480;	DB 12;	Length 520;
Best Local Similarity	91.9%;	Pred. No. 7.3e-238;		
Matches 478;	Conservative	0;	Mismatches 0;	Indels 42;
				Gaps 1
QY	1	MPGSLPLNAAECWPKDVGIVALEITYEPPSQYVDQAELEKXKVDGADGAKYTTIGLQAKMGFCT	60	
Db	1	MPGSLPLNAAECWPKDVGIVALEITYEPPSQYVDQAELEKXKVDGADGAKYTTIGLQAKMGFCT	60	
QY	61	DREDINSLCMTFVQNLMERNNLSYDCIGRLEVGTEETIIDKSKSVKTNLMQLEESGNTDI	120	
Db	61	DREDINSLCMTFVQNLMERNNLSYDCIGRLEVGTEETIIDKSKSVKTNLMQLEESGNTDI	120	
QY	121	EGIDITNACYGCGTAAVFNAAVNMIESSWD-----	149	
Db	121	EGIDITNACYGCGTAAVFNAAVNMIESSWDGRYALVVAQDIAVYATGNARPTGCVGAVALL	180	
QY	150	-----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLISIQCYLSALDRCSVYCCKI	198	
Db	181	IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLISIQCYLSALDRCSVYCCKI	240	
QY	199	HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDXNSIYSGLEA	258	
Db	241	HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDXNSIYSGLEA	300	
QY	259	EGDVKLEDTYFPDRVEKAFMKASSELESQTKASILVSNQNGNMNTSSVYGSILASVLAQY	318	

Db 301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTASLVSNGNMYSVYGSGLASVLAQY 360
QY 319 SPOQLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGYAP 378
Db 361 SPOQLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGYAP 420
QY 379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520

RESULT 4
US-10-236-417-184

; Sequence 184, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: CuiCom
; SEQ ID NO 184
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-417-184

Query Match 98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYTTIGLGAQKMGFCT 60
Db 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYTTIGLGAQKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNLNSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNLNSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY 121 EGIPTTNACYGCTAAVFNANWNISSWD----- 149
Db 121 EGIPTTNACYGCTAAVFNANWNISSWDGRYALVVAAGDIAYATGNARPTGCVGAVALL 180
QY 150 -----GLRGTHMOAHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
Db 150 -----GLRGTHMOAHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198

Db 181 IGBNAPLIFERGLRGTHMOAHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240
QY 199 HAQWQKEGNDKDFTLNDFGEMI FHSPLYCKLVOKSLARMNLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQKEGNDKDFTLNDFGEMI FHSPLYCKLVOKSLARMNLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTASLVSNGNMYSVYGSGLASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTASLVSNGNMYSVYGSGLASVLAQY 360
QY 319 SPOQLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGYAP 378
Db 361 SPOQLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGYAP 420
QY 379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520

RESULT 5
US-10-236-417-186

; Sequence 186, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: CuiCom
; SEQ ID NO 186
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-417-186

Query Match 98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYTTIGLGAQKMGFCT 60
Db 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYTTIGLGAQKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNLNSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNLNSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

[illegible]

```

RESULT 6
US-10-236-417-190
; Sequence 190, Application US//10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agsee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 190
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-417-190

```

Query Match	98.8%;	Score 2480;	DB 12;	Length 520;
Best Local Similarity	91.9%;	Pred. No. 7.3e-238;		
Matches 478; Conservative	0;	Mismatches 0;	Indels 42;	Gaps 1;

QY	1	MPGSLPLNAEACWPKDVGIVALEIYFSPQYVDQAELEKXDGVADGKTYITGLGQAKMGFCT	60
Db	1	MPGSLPLNAEACWPKDVGIVALEIYFSPQYVDQAELEKXDGVADGKTYITGLGQAKMGFCT	60
QY	61	DREDINSLCMTVVQNMERNNLSDYCIGRLEVGTEITIIDKSKSVKTNLMQLFEESGNTDI	120
Db	61	DREDINSLCMTVVQNMERNNLSDYCIGRLEVGTEITIIDKSKSVKTNLMQLFEESGNTDI	120
QY	121	EGIDTTNACYGGTAAVFNAVNMISSGMD-----	149
Db	121	EGIDTTNACYGGTAAVFNAVNMISSGMDGRYALVAVAGDIAYATAGNARPTGCGAVALL	180
QY	150	-----GLRGTHMÔHAYDFYKPKDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKI	198
Db	181	IGNAPLIFERGLRGTHMÔHAYDFYKPKDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKI	240
QY	199	HAOMQKEGNDKDFTLNDFGEMI FHS PYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA	258
Db	241	HAOMQKEGNDKDFTLNDFGEMI FHS PYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA	300
QY	259	FGDYKLEDTYFDRDVEKA FMKASSELSÔKTKASLLVSNONGNMYTSSVYGSILASVLAQY	318
Db	301	FGDYKLEDTYFDRDVEKA FMKASSELSÔKTKASLLVSNONGNMYTSSVYGSILASVLAQY	360
QY	319	SPQOLAGKRIGVFSYSGSLAATLYSLKYTQDATPGSALDKTTASLCDLKSRLDSRTGVAP	378
Db	361	SPQOLAGKRIGVFSYSGSLAATLYSLKYTQDATPGSALDKTTASLCDLKSRLDSRTGVAP	420
QY	379	DVFAENMKLREDTTHLVNYI PÔGSIDSLEFGTWYLVVRVDEKRRTYARRPTPNDDTLDEG	438
Db	421	DVFAENMKLREDTTHLVNYI PÔGSIDSLEFGTWYLVVRVDEKRRTYARRPTPNDDTLDEG	480
QY	439	VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478	
Db	481	VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520	

```

RESULT 7
US-10-236-417-194
; Sequence 194, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 194
; LENGTH: 520
; TYPE: PRT
;

```

```
; ORGANISM: Homo sapiens
US-10-236-417-194

Query Match          98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY      1  MEGSLPLNAAECWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60
      1  MEGSLPLNAAECWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60
DB      1  MEGSLPLNAAECWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60

QY      61  DREDINSLCMTVVQNLMERNNLSDYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
      61  DREDINSLCMTVVQNLMERNNLSDYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
DB      61  DREDINSLCMTVVQNLMERNNLSDYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

QY      121  EGIDTTCACGCGTAAFNAVNWIESSSWD----- 149
      121  EGIDTTCACGCGTAAFNAVNWIESSSWD----- 149
DB      121  EGIDTTCACGCGTAAFNAVNWIESSSWDGRYALVVGADIAVYATGNARPTGGVAVALL 180

QY      150  -----GLRGTHMQHAYDFYKPPDMLSEYPIVDGKLSIQCYLSALDRCSYVCCKI 198
      150  -----GLRGTHMQHAYDFYKPPDMLSEYPIVDGKLSIQCYLSALDRCSYVCCKI 198
DB      181  IGPNAPLIFERGLRGTHMQHAYDFYKPPDMLSEYPIVDGKLSIQCYLSALDRCSYVCCKI 240

QY      199  HAOWOKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA 258
      199  HAOWOKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA 258
DB      241  HAOWOKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA 300

QY      259  FGDVPLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
      259  FGDVPLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
DB      301  FGDVPLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

QY      319  SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
      319  SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
DB      361  SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

QY      379  DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWTYLVRVDEKGRRTYARPTPNDDTLDEG 438
      379  DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWTYLVRVDEKGRRTYARPTPNDDTLDEG 438
DB      421  DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWTYLVRVDEKGRRTYARPTPNDDTLDEG 480

QY      439  VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
      439  VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
DB      481  VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520

RESULT 8
US-10-236-417-198
; Sequence 198, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
```

```
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 198
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-417-198

Query Match          98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY      1  MEGSLPLNAAECWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60
      1  MEGSLPLNAAECWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60
DB      1  MEGSLPLNAAECWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60

QY      61  DREDINSLCMTVVQNLMERNNLSDYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
      61  DREDINSLCMTVVQNLMERNNLSDYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
DB      61  DREDINSLCMTVVQNLMERNNLSDYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

QY      121  EGIDTTCACGCGTAAFNAVNWIESSSWD----- 149
      121  EGIDTTCACGCGTAAFNAVNWIESSSWD----- 149
DB      121  EGIDTTCACGCGTAAFNAVNWIESSSWDGRYALVVGADIAVYATGNARPTGGVAVALL 180

QY      150  -----GLRGTHMQHAYDFYKPPDMLSEYPIVDGKLSIQCYLSALDRCSYVCCKI 198
      150  -----GLRGTHMQHAYDFYKPPDMLSEYPIVDGKLSIQCYLSALDRCSYVCCKI 198
DB      181  IGPNAPLIFERGLRGTHMQHAYDFYKPPDMLSEYPIVDGKLSIQCYLSALDRCSYVCCKI 240

QY      199  HAOWOKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA 258
      199  HAOWOKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA 258
DB      241  HAOWOKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA 300

QY      259  FGDVPLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
      259  FGDVPLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
DB      301  FGDVPLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

QY      319  SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
      319  SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
DB      361  SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

QY      379  DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWTYLVRVDEKGRRTYARPTPNDDTLDEG 438
      379  DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWTYLVRVDEKGRRTYARPTPNDDTLDEG 438
DB      421  DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWTYLVRVDEKGRRTYARPTPNDDTLDEG 480

QY      439  VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
      439  VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
DB      481  VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520

RESULT 9
US-10-236-417-202
; Sequence 202, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
```

```
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 202
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-417-202
```

```
Query Match          98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
```

```
QY      1 MPGLPLNABACWPKDVGIVALEIYFPSQYVDQAELEKYGVDAGKYYTIGLQAKMGFCT 60
        |||||||
DB       1 MPGLPLNABACWPKDVGIVALEIYFPSQYVDQAELEKYGVDAGKYYTIGLQAKMGFCT 60
QY      61 DREDINSLCMTVVQNLMERNNLSYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
        |||||||
DB       61 DREDINSLCMTVVQNLMERNNLSYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY      121 EGIDTTNACYGCGTAAVNAVNWIESSWD----- 149
        |||||||
DB       121 EGIDTTNACYGCGTAAVNAVNWIESSWDGRYALVVAAGDIAVYATGNARPTGCVGAVALL 180
QY      150 -----GLRGTHNQHA YDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
        |||||||
DB       181 IGPNAPLIFERGLRGTHNQHA YDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240
QY      199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
        |||||||
DB       241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY      259 FGDVKLEDTYFDRDVEKAFMKASSELFSQTKASLLVSNQNGNMTSSVYGSLSAVLAQY 318
        |||||||
DB       301 FGDVKLEDTYFDRDVEKAFMKASSELFSQTKASLLVSNQNGNMTSSVYGSLSAVLAQY 360
QY      319 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
        |||||||
DB       361 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY      379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEG 438
        |||||||
DB       421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEG 480
QY      439 VGLVHSNIATEHIIPSPAKKYVRLPATAAEPAAVISNGEH 478
        |||||||
DB       481 VGLVHSNIATEHIIPSPAKKYVRLPATAAEPAAVISNGEH 520
```

```
RESULT 10
US-10-236-417-204
; Sequence 204, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
```

```
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 204
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-417-204
```

```
Query Match          98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
```

```
QY      1 MPGLPLNABACWPKDVGIVALEIYFPSQYVDQAELEKYGVDAGKYYTIGLQAKMGFCT 60
        |||||||
DB       1 MPGLPLNABACWPKDVGIVALEIYFPSQYVDQAELEKYGVDAGKYYTIGLQAKMGFCT 60
QY      61 DREDINSLCMTVVQNLMERNNLSYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
        |||||||
DB       61 DREDINSLCMTVVQNLMERNNLSYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY      121 EGIDTTNACYGCGTAAVNAVNWIESSWD----- 149
        |||||||
DB       121 EGIDTTNACYGCGTAAVNAVNWIESSWDGRYALVVAAGDIAVYATGNARPTGCVGAVALL 180
QY      150 -----GLRGTHNQHA YDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
        |||||||
DB       181 IGPNAPLIFERGLRGTHNQHA YDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240
QY      199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
        |||||||
DB       241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY      259 FGDVKLEDTYFDRDVEKAFMKASSELFSQTKASLLVSNQNGNMTSSVYGSLSAVLAQY 318
        |||||||
DB       301 FGDVKLEDTYFDRDVEKAFMKASSELFSQTKASLLVSNQNGNMTSSVYGSLSAVLAQY 360
QY      319 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
        |||||||
DB       361 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY      379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEG 438
        |||||||
DB       421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEG 480
QY      439 VGLVHSNIATEHIIPSPAKKYVRLPATAAEPAAVISNGEH 478
        |||||||
DB       481 VGLVHSNIATEHIIPSPAKKYVRLPATAAEPAAVISNGEH 520
```

```
RESULT 11
US-10-236-417-206
; Sequence 206, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
```



```
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 206
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-417-206
```

```
Query Match          98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
```

```
QY      1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVADGKTTIGLGQAKMGFCT 60
      |||||||
Db       1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVADGKTTIGLGQAKMGFCT 60

QY      61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMOLFEESENTDI 120
      |||||||
Db       61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMOLFEESENTDI 120

QY      121 EGIDTTNACYGCTAAVFNAVNWIESSSWD----- 149
      |||||||
Db       121 EGIDTTNACYGCTAAVFNAVNWIESSSWDGRYALVVAGDIAYATGNARPTGVGAVALL 180

QY      150 -----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
      |||||||
Db       181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

QY      199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
      |||||||
Db       241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300

QY      259 FGDVYKLEDITYPDRDVEKAFMKASSELSFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQY 318
      |||||||
Db       301 FGDVYKLEDITYPDRDVEKAFMKASSELSFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQY 360

QY      319 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
      |||||||
Db       361 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420

QY      379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDEG 438
      |||||||
Db       421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDEG 480

QY      439 VGLVHNSNIAATEHISPAPKAVPRLPATAAEPEAAVISNGEH 478
      |||||||
Db       481 VGLVHNSNIAATEHISPAPKAVPRLPATAAEPEAAVISNGEH 520
```

RESULT 12

```
US-10-236-417-208
; Sequence 208, Application US/10236417
; Publication No. US20040048256a1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 208
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-417-208
```

```
Query Match          98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
```

```
QY      1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVADGKTTIGLGQAKMGFCT 60
      |||||||
Db       1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVADGKTTIGLGQAKMGFCT 60

QY      61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMOLFEESENTDI 120
      |||||||
Db       61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMOLFEESENTDI 120

QY      121 EGIDTTNACYGCTAAVFNAVNWIESSSWD----- 149
      |||||||
Db       121 EGIDTTNACYGCTAAVFNAVNWIESSSWDGRYALVVAGDIAYATGNARPTGVGAVALL 180

QY      150 -----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
      |||||||
Db       181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

QY      199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
      |||||||
Db       241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300

QY      259 FGDVYKLEDITYPDRDVEKAFMKASSELSFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQY 318
      |||||||
Db       301 FGDVYKLEDITYPDRDVEKAFMKASSELSFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQY 360

QY      319 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
      |||||||
Db       361 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420

QY      379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDEG 438
      |||||||
Db       421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDEG 480
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2004, 13:16:22 ; Search time 49 Seconds
(without alignments)
503.617 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPGLPLNAEACWPXDVGVIV.....PRLPATAEPEAAVISNGEH 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2511	100.0	478	4	US-09-819-993-2 Sequence 2, Appli
2	2511	100.0	478	4	US-10-193-295-2 Sequence 2, Appli
3	2480	98.8	520	4	US-09-819-993-4 Sequence 4, Appli
4	2480	98.8	520	4	US-10-193-295-4 Sequence 4, Appli
5	2439	97.1	518	4	US-09-819-993-5 Sequence 5, Appli
6	2439	97.1	518	4	US-10-193-295-5 Sequence 5, Appli
7	2439	97.1	520	1	US-08-305-505-6 Sequence 6, Appli
8	2362	94.1	520	1	US-08-305-505-5 Sequence 5, Appli
9	2352	93.7	520	1	US-08-305-505-4 Sequence 4, Appli
10	2097	83.5	522	1	US-08-305-505-2 Sequence 2, Appli
11	1476.5	58.8	507	1	US-08-305-505-3 Sequence 3, Appli
12	912.5	36.3	467	3	US-09-306-595C-6 Sequence 6, Appli
13	912.5	36.3	467	4	US-09-925-388-6 Sequence 6, Appli
14	686	27.3	168	4	US-09-401-064-198 Sequence 198, App
15	315	12.5	388	4	US-09-107-532A-5369 Sequence 5369, Ap
16	284.5	11.3	436	4	US-09-01C-4295 Sequence 4295, Ap
17	282	11.2	449	4	US-09-134-000C-5444 Sequence 5444, Ap
18	209.5	8.3	316	4	US-09-241-750-2 Sequence 2, Appli
19	123.5	4.9	1495	4	US-09-543-681A-5986 Sequence 5986, Ap
20	123	4.9	358	4	US-09-543-681A-8134 Sequence 8134, Ap
21	105.5	4.2	1143	3	US-08-310-912A-108 Sequence 108, App
22	105.5	4.2	1143	2	US-09-301-085-108 Sequence 108, App
23	105.5	4.2	1143	5	PCT-US95-04589-108 Sequence 108, App
24	105.5	4.2	1144	1	US-08-261-663A-2 Sequence 2, Appli
25	105.5	4.2	1144	1	US-08-261-663A-4 Sequence 4, Appli
26	105.5	4.2	1144	4	US-09-357-206A-3 Sequence 3, Appli
27	105.5	4.2	1144	4	US-09-813-742A-3 Sequence 3, Appli

28	105.5	4.2	1144	5	PCT-US95-07754A-2	Sequence 2, Appli
29	105.5	4.2	1144	5	PCT-US95-07754A-4	Sequence 4, Appli
30	105	4.2	1494	3	US-08-755-587-186	Sequence 186, App
31	103.5	4.1	3169	4	US-09-453-702B-257	Sequence 257, App
32	103	4.1	1252	4	US-10-012-762-20	Sequence 20, Appl
33	103	4.1	1252	4	US-09-704-036B-20	Sequence 20, Appl
34	102.5	4.1	1144	3	US-08-930-996A-9	Sequence 9, Appli
35	102	4.1	784	3	US-08-846-234-5	Sequence 5, Appli
36	99.5	4.0	652	1	US-08-261-663A-6	Sequence 6, Appli
37	99.5	4.0	652	4	US-09-357-206A-5	Sequence 5, Appli
38	99.5	4.0	652	4	US-09-813-742A-5	Sequence 5, Appli
39	99.5	4.0	652	5	PCT-US95-07754A-6	Sequence 6, Appli
40	99.5	4.0	876	1	US-08-785-429-2	Sequence 2, Appli
41	99.5	4.0	876	3	US-08-996-621-2	Sequence 2, Appli
42	96.5	3.8	616	4	US-09-236-063-1	Sequence 1, Appli
43	96.5	3.8	640	4	US-09-351-814-13	Sequence 13, Appl
44	96.5	3.8	3248	1	US-08-353-700-1	Sequence 1, Appli
45	96.5	3.8	3248	5	PCT-US95-16216-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-819-993-2
; Sequence 2, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819, 993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
; US-09-819-993-2

Query Match	Score	DB 4;	Length	478;
Best Local Similarity	100.0%;	Pred. No. 1.9e-245;		
Matches	478;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MPGLPLNAEACWPXDVGVIVALEIYFPSQYVDQAELEKYDGDVDAKTYTIGLGQAKMGFCT	60	
Db	1	MPGLPLNAEACWPXDVGVIVALEIYFPSQYVDQAELEKYDGDVDAKTYTIGLGQAKMGFCT	60	
QY	61	DREDINSLCMTVQNLMERNNLSYDCIGRLVEGTETIIDKSKSVKTNLMQLFEESGNTDI	120	
Db	61	DREDINSLCMTVQNLMERNNLSYDCIGRLVEGTETIIDKSKSVKTNLMQLFEESGNTDI	120	
QY	121	EGIDITNACYGGAFFNAVNWIIESSSWDGLRGTHMOHAYDYFKPDMLSPIYDGKLSI	180	
Db	121	EGIDITNACYGGAFFNAVNWIIESSSWDGLRGTHMOHAYDYFKPDMLSPIYDGKLSI	180	
QY	181	QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND	240	
Db	181	QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND	240	
QY	241	FLNDQNRDKNSIYSGIAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLIVSNQNG	300	
Db	241	FLNDQNRDKNSIYSGIAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLIVSNQNG	300	
QY	301	NMYTSSVYGSILASVLAQYSPQQLAGKRIGVFSGSLAATLYSLKVTQDATPSPALDKIT	360	
Db	301	NMYTSSVYGSILASVLAQYSPQQLAGKRIGVFSGSLAATLYSLKVTQDATPSPALDKIT	360	
QY	361	ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHLVNYIPQGSIDSLFEGTWTLYRVDEKH	420	

Db 361 ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH 420
QY 421 RRTYARRPTPNDDTLDEGVGLVHSHNATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
421 RRTYARRPTPNDDTLDEGVGLVHSHNATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db

RESULT 2
US-10-193-295-2
; Sequence 2, Application US/10193295
; Patent No. 6620608
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195DIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
US-10-193-295-2

Query Match 100.0%; Score 2511; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No.1.9e-245;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPGSLPLNAEACWPKDVGIIVALEIYFSPQYVDQAELEKXDVDAKXTTIGLGOAKMGFCT 60
1 MPGSLPLNAEACWPKDVGIIVALEIYFSPQYVDQAELEKXDVDAKXTTIGLGOAKMGFCT 60
Db
QY 61 DREDINSLCMTVVQNMERNNLSYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
61 DREDINSLCMTVVQNMERNNLSYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db
QY 121 EGIDTTNACYGGTAAVFNAVNWIESSSDGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSI 180
121 EGIDTTNACYGGTAAVFNAVNWIESSSDGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSI 180
Db
QY 181 QCYLSALDRCYSVYCKKTHAQWQKEGNDKFTLNDFGFMI FHSPLYCKLVQKSLARMLND 240
181 QCYLSALDRCYSVYCKKTHAQWQKEGNDKFTLNDFGFMI FHSPLYCKLVQKSLARMLND 240
Db
QY 241 FLNDQNRDKNSIYSGLEAFGVDKLEDYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG 300
241 FLNDQNRDKNSIYSGLEAFGVDKLEDYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG 300
Db
QY 301 NMYTSSVYGSGLASVLAQYSPQOLAGKRIGVFSYSGSLAATLYSLKVTQDAPPGSALDKIT 360
301 NMYTSSVYGSGLASVLAQYSPQOLAGKRIGVFSYSGSLAATLYSLKVTQDAPPGSALDKIT 360
Db
QY 361 ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH 420
361 ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH 420
Db
QY 421 RRTYARRPTPNDDTLDEGVGLVHSHNATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
421 RRTYARRPTPNDDTLDEGVGLVHSHNATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db

RESULT 3
US-09-819-993-4
; Sequence 4, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Human
US-09-819-993-4

Query Match 98.8%; Score 2480; DB 4; Length 520;
Best Local Similarity 91.9%; Pred. No.3.1e-242;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MPGSLPLNAEACWPKDVGIIVALEIYFSPQYVDQAELEKXDVDAKXTTIGLGOAKMGFCT 60
1 MPGSLPLNAEACWPKDVGIIVALEIYFSPQYVDQAELEKXDVDAKXTTIGLGOAKMGFCT 60
Db
QY 61 DREDINSLCMTVVQNMERNNLSYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
61 DREDINSLCMTVVQNMERNNLSYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db
QY 121 EGIDTTNACYGGTAAVFNAVNWIESSSDGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSI 149
121 EGIDTTNACYGGTAAVFNAVNWIESSSDGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSI 149
Db
QY 150 -----GLRGTHMQHAYDFYKPDMLSEYPVVDGKLSIQCYLALDRCYSVYCKKI 198
150 -----GLRGTHMQHAYDFYKPDMLSEYPVVDGKLSIQCYLALDRCYSVYCKKI 198
Db
QY 181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSIQCYLSALDRCYSVYCKKI 240
181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSIQCYLSALDRCYSVYCKKI 240
Db
QY 199 HAQWQKEGNDKFTLNDFGFMI FHSPLYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
199 HAQWQKEGNDKFTLNDFGFMI FHSPLYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db
QY 241 HAQWQKEGNDKFTLNDFGFMI FHSPLYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
241 HAQWQKEGNDKFTLNDFGFMI FHSPLYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
Db
QY 259 FGDVLEDYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQY 318
259 FGDVLEDYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQY 318
Db
QY 301 FGDVLEDYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQY 360
301 FGDVLEDYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQY 360
Db
QY 319 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDAPPGSALDKITASLCDLKSRLDSRTGVAP 378
319 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDAPPGSALDKITASLCDLKSRLDSRTGVAP 378
Db
QY 361 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDAPPGSALDKITASLCDLKSRLDSRTGVAP 420
361 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDAPPGSALDKITASLCDLKSRLDSRTGVAP 420
Db
QY 379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRTYARRPTPNDDTLDEG 438
379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRTYARRPTPNDDTLDEG 438
Db
QY 421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRTYARRPTPNDDTLDEG 480
421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRTYARRPTPNDDTLDEG 480
Db
QY 439 VGLVHSHNATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
439 VGLVHSHNATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db
QY 481 VGLVHSHNATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520
481 VGLVHSHNATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520
Db

RESULT 4
US-10-193-295-4
; Sequence 4, Application US/10193295
; Patent No. 6620608
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195DIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 520
; TYPE: PRT


```

; ORGANISM: Human
US-10-193-295-4

Query Match      98.8%; Score 2480; DB 4; Length 520;
Best Local Similarity 91.9%; Pred. No. 3.1e-242;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60
DB      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60

QY      61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
DB      61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

QY      121 EGDITTNACYGGTAAVFNAVNWIESSSWDGRYALVAGDIAVYATGNARPTGCGAVALL 180
DB      121 EGDITTNACYGGTAAVFNAVNWIESSSWDGRYALVAGDIAVYATGNARPTGCGAVALL 180

QY      150 -----GLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
DB      181 IGPNAPLIFERGLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

QY      199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
DB      241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300

QY      259 FGDVYKLEDITYFDRDVEKAFMKASSELFSQKTKASLTVSNQNGNMYTSSVYGSILASVLAQY 318
DB      301 FGDVYKLEDITYFDRDVEKAFMKASSELFSQKTKASLTVSNQNGNMYTSSVYGSILASVLAQY 360

QY      319 SPOQLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
DB      361 SPOQLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420

QY      379 DVFAENMKLREDTHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 438
DB      421 DVFAENMKLREDTHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 480

QY      439 VGLVHSNIATEHIIPSPAKKVPRLPATAAPEAAVISNGEH 478
DB      481 VGLVHSNIATEHIIPSPAKKVPRLPATAAPEAAVISNGEH 520

RESULT 5
US-09-819-993-5
; Sequence 5, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Human
US-09-819-993-5

Query Match      97.1%; Score 2439; DB 4; Length 518;
Best Local Similarity 90.9%; Pred. No. 4.3e-238;
Matches 471; Conservative 1; Mismatches 4; Indels 42; Gaps 1;

QY      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60
DB      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60

QY      61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
DB      61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

QY      121 EGDITTNACYGGTAAVFNAVNWIESSSWDGRYALVAGDIAVYATGNARPTGCGAVALL 180
DB      121 EGDITTNACYGGTAAVFNAVNWIESSSWDGRYALVAGDIAVYATGNARPTGCGAVALL 180

QY      150 -----GLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
DB      181 IGPNAPLIFERGLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240
```

```

DB      61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY      121 EGDITTNACYGGTAAVFNAVNWIESSSWD----- 149
DB      121 EGDITTNACYGGTAAVFNAVNWIESSSWDGRYALVAGDIAVYATGNARPTGCGAVALL 180
QY      150 -----GLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
DB      181 IGPNAPLIFERGLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

QY      199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
DB      241 HAQWQKEANDNDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300

QY      259 FGDVYKLEDITYFDRDVEKAFMKASSELFSQKTKASLTVSNQNGNMYTSSVYGSILASVLAQY 318
DB      301 FGDVYKLEDITYFDRDVEKAFMKASSELFSQKTKASLTVSNQNGNMYTSSVYGSILASVLAQY 360

QY      319 SPOQLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
DB      361 SPOHLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAQ 420

QY      379 DVFAENMKLREDTHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 438
DB      421 DVFAENMKLREDTHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 480

QY      439 VGLVHSNIATEHIIPSPAKKVPRLPATAAPEAAVISNG 476
DB      481 VGLVHSNIATEHIIPSPAKKVPRLPATAAPEAAVISNG 518

RESULT 6
US-10-193-295-5
; Sequence 5, Application US/10193295
; Patent No. 6620608
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195DIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Human
US-10-193-295-5

Query Match      97.1%; Score 2439; DB 4; Length 518;
Best Local Similarity 90.9%; Pred. No. 4.3e-238;
Matches 471; Conservative 1; Mismatches 4; Indels 42; Gaps 1;

QY      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60
DB      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTTIGLGQAKMGFCT 60

QY      61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
DB      61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

QY      121 EGDITTNACYGGTAAVFNAVNWIESSSWD----- 149
DB      121 EGDITTNACYGGTAAVFNAVNWIESSSWDGRYALVAGDIAVYATGNARPTGCGAVALL 180

QY      150 -----GLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
DB      181 IGPNAPLIFERGLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240
```

QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQKEANDNDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDKLEDTYFDRDVEKAFMKASSELFSQTKASLLVSNQNGNMYSVYGSILASVLAQY 318
Db 301 FGDKLEDTYFDRDVEKAFMKASSELFSQTKASLLVSNQNGNMYSVYGSILASVLAQY 360
QY 319 SPQOLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db 361 SPQOLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVEAENMKLRDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDEG 438
Db 421 DVEAENMKLRDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIIPSPAKKVPRLPATAEPEAAVISNG 476
Db 481 VGLVHSNIATEHIIPSPAKKVPRLPATAEPEAAVISNG 518

RESULT 7
US-08-305-505-6
; Sequence 6, Application US/08305505
; Patent No. 5668001
; GENERAL INFORMATION:
; APPLICANT: Miziorko, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
; TITLE OF INVENTION: STABILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,040
; FILING DATE: 02 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 65-053-9083-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 520 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-305-505-6

Query Match 97.1%; Score 2439; DB 1; Length 520;
Best Local Similarity 90.9%; Pred. No. 4.4e-238;
Matches 471; Conservative 1; Mismatches 4; Indels 42; Gaps 1;

QY 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKXDGVDAGKYITIGLGQAKMGFCT 60
|||||

Db 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKXDGVDAGKYITIGLGQAKMGFCT 60
QY 61 DREDINSLCMTYVQNLMERNNLSYDCIGRLEVGTEITIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTYVQNLMERNNLSYDCIGRLEVGTEITIDKSKSVKTNLMQLFEESGNTDI 120
QY 121 EGI DTNACYGGTA AVENAVNMTESSSWD----- 149
Db 121 EGI DTNACYGGTA AVENAVNMTESSSWDGRYALVWAGDIAYATGNARPTGGVAVALL 180
QY 150 -----GLRGTMOHAYDFYKPDMLSEYPIVDGKLSTQCYSALDRCSYCYCKI 198
Db 181 IGBNAPLIFERGLRGTMOHAYDFYKPDMLSEYPIVDGKLSTQCYSALDRCSYCYCKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQKEANDNDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDKLEDTYFDRDVEKAFMKASSELFSQTKASLLVSNQNGNMYSVYGSILASVLAQY 318
Db 301 FGDKLEDTYFDRDVEKAFMKASSELFSQTKASLLVSNQNGNMYSVYGSILASVLAQY 360
QY 319 SPQOLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db 361 SPQOLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVEAENMKLRDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDEG 438
Db 421 DVEAENMKLRDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIIPSPAKKVPRLPATAEPEAAVISNG 476
Db 481 VGLVHSNIATEHIIPSPAKKVPRLPATAEPEAAVISNG 518

RESULT 8
US-08-305-505-5
; Sequence 5, Application US/08305505
; Patent No. 5668001
; GENERAL INFORMATION:
; APPLICANT: Miziorko, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
; TITLE OF INVENTION: STABILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,040
; FILING DATE: 02 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 65-053-9083-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

LENGTH: 520 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-305-505-5

Query Match 94.1%; Score 2362; DB 1; Length 520;
Best Local Similarity 87.5%; Pred No. 2.8e-230;
Matches 455; Conservative 11; Mismatches 12; Indels 42; Gaps 1;

QY 1 MEGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60
Db 1 MEGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQARMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCLTVVQKLMERNLSYDCIGRLEVGTETIIDKSKSVKSNLMQLFEESGNTDI 120
QY 121 EGIDTTNACYGGTAAVFNAVNWIESSSWD----- 149
Db 121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVAGDIAIYASGNARPTGGVGAVALL 180
QY 150 -----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
Db 181 IGPNAFVIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSIQCYLSALDRCYSVYRKKI 240
QY 199 HAOWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 258
Db 241 RAQWQKEGKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVYKLEDYFDRDVEKAFMKASSELFSQKTASLVSNOGNMYTSSVYGSILASVLAQY 318
Db 301 FGDVYKLEDYFDRDVEKAFMKASAELENQKTASLVSNOGNMYTSSVYGSILASVLAQY 360
QY 319 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db 361 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVFAENMKLREDTHHLVNIIPQCSIDSLFEGTWYLVRVDEKGRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLVNIIPQCSIDSLFEGTWYLVRVDEKGRRTYARRPTSTNDHSLDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNTATEHIPSPAKKVPRLPATSGEPESAVISNGEH 520

RESULT 9
US-08-305-505-4
Sequence 4, Application US/08305505
Patent No. 5668001
GENERAL INFORMATION:
APPLICANT: Miziorko, Henry M.
TITLE OF INVENTION: 3-HYDROXY-3-METHYLGUTARYL-CoA
TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
TITLE OF INVENTION: STABILITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/305,505
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/072,040
FILING DATE: 02 JUNE 1993
ATTORNEY/AGENT INFORMATION:
NAME: Baker, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 65-053-9083-9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5709
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 520 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-305-505-4

Query Match 93.7%; Score 2352; DB 1; Length 520;
Best Local Similarity 87.5%; Pred. No. 2.8e-229;
Matches 455; Conservative 9; Mismatches 14; Indels 42; Gaps 1;

QY 1 MEGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60
Db 1 MEGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQARMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCLTVVQKLMERNLSYDCIGRLEVGTETIIDKSKSVKSNLMQLFEESGNTDI 120
QY 121 EGIDTTNACYGGTAAVFNAVNWIESSSWD----- 149
Db 121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVAGDIAIYATGNARPTGGVGAVALL 180
QY 150 -----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
Db 181 IGPNAFLIFDRGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYRKKI 240
QY 199 HAOWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 258
Db 241 RAQWQKEGNDNDFTLNDFGFMISHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVYKLEDYFDRDVEKAFMKASSELFSQKTASLVSNOGNMYTSSVYGSILASVLAQY 318
Db 301 FGDVYKLEDYFDRDVEKAFMKASSELFNQKTASLVSNOGNMYTSSVYGSILASVLAQY 360
QY 319 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db 361 SPQOLAGKRIGVFSYSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVFAENMKLREDTHHLVNIIPQCSIDSLFEGTWYLVRVDEKGRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLVNIIPQCSIDSLFEGTWYLVRVDEKGRRTYARRPTSTNDHNLGDG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNTATEHIPSPAKKVPRLPATAESESAVISNGEH 520

RESULT 10
US-08-305-505-2
Sequence 2, Application US/08305505
Patent No. 5668001
GENERAL INFORMATION:
APPLICANT: Miziorko, Henry M.
TITLE OF INVENTION: 3-HYDROXY-3-METHYLGUTARYL-CoA
TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
TITLE OF INVENTION: STABILITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady


```

; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,040
; FILING DATE: 02 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 65-053-9083-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-305-505-2

```

Query Match 83.5%; Score 2097; DB 1; Length 522;
Best Local Similarity 76.6%; Pred. No. 2e-203;
Matches 400; Conservative 36; Mismatches 42; Indels 44; Gaps 3;

```

QY 1 MPGSLPLNBAECWPKDVGIIVALEIYFPPSQYVDQAELEKYDGVADGKYTIGLGOAKMGFCT 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MPGSLPVNTESCMPKDVGIIVALEIYFPPSQYVDQTELEKYDGVADGKYTIGLGOAKMGFCS 60
QY 61 DREDINSLCMTVVQNLMEERNLSYDCIGRLVETETIIDKSKSVKTNLMQLFEESGNTDI 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DREDINSLCMTVVQNLMEERNLSYDCIGRLVETETIIDKSKSVKTNLMQLFEESGNTDV 120
QY 121 EGIDTTNACYGTAAVFNAVNWIESSWD----- 149
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 EGIDTTNACYGTAALFNAINWIESSSWDGRYALVAGDIAYATGNARPTGAGAVAML 180
QY 150 -----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCTSVYCKKI 198
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 VGSNAPLIFERGLRGTHMQHAYDFYKPDWSEYFVVDGKLSIQCYLSALDRCTSVYRNKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMI FHSPYCKLVQKSLARMLNDFLNDQNRD-KNSIYSGLE 257
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 HAQWQKEGTRGF TLNDFGFMI FHSPYCKLVQKSVARLLNDFLSDQNAETANGVFSGLE 300
QY 258 AFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQ 317
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 AFRDVKLEDTYFDRDVEKAFMKASAE LFNQTKASLLVSNQNGNMYTPSVYGC LASVLAQ 360
QY 318 YSPQQLAGKRIGVFSYSGSLATLYSLKVTODATPGSALDKITASLCDLKSRLDSRTGVA 377
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 YSPEHLAGQRISEFSYSGSFAATLYSIRVTDATPGSALDKITASLSDLKARLDSRKCI A 420
QY 378 PDVFAENMKLRBDTHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTFNDTLD E 437
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 PDVFAENMKIRQETHLANYPQCSVEDLFEGTWYLVRVDEKRRRTYARRRPWG DGPL E A 480
QY 438 GVLVHNSIATEHIPSAPKAVPRLPATA-AEPEAAVISNGEH 478
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 GVEVVAHPGIVHEHIPSAPKAVPRIPATTESEGVTAISNGVH 522

```

```

RESULT 11
US-08-305-505-3
; Sequence 3, Application US/08305505
; Patent No. 5668001
;
; GENERAL INFORMATION:
; APPLICANT: Miziozko, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGUTARYL-CoA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
; TITLE OF INVENTION: STABILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,040
; FILING DATE: 02 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 65-053-9083-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 507 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-305-505-3

```

Query Match 58.8%; Score 1476.5; DB 1; Length 507;
Best Local Similarity 58.9%; Pred. No. 1.4e-140;
Matches 274; Conservative 76; Mismatches 72; Indels 43; Gaps 2;

```

QY 6 PLNBAECWPKDVGIIVALEIYFPPSQYVDQAELEKYDGVADGKYTIGLGOAKMGFCTDREDI 65
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 43 PLAKTDTWPKDVGIIVALEIYFPPAQYVDQTELEKFNNV EAGKYTVGLGQTRMGFCSVQEDI 102
QY 66 NSLCMTVVQNLMEERNLSYDCIGRLVETETIIDKSKSVKTNLMQLFEESGNTDIEGIDT 125
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 103 NSLCITVVQRLMERKLPWDHVRGLVETETIIDKSKAVKTYLMELFQDSGNTDIEGIDT 162
QY 126 TNACYGGTAAVFNAVNWIESSWD----- 149
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 TNACYGGTASLFNAANMMESSYWDGRYALVVGCDIAYVPSGNRPTGAGAVAM LIGPKA 222
QY 150 -----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCTSVYCKKIHAQWQ 203
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 223 PLVLEQGLRGTHMENAYDFYKPNLASEYPIVDGKLSIQCYLRALDRCYAAYRRKIQNQWK 282
QY 204 KEGNDKDFTLNDFGFMI FHSPYCKLVQKSLARMLNDFLNDQNRDKNISIYSGLEAFGDVK 263
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 283 QAGNNQPFITLDVQYMI FHTPRCKMVQKSLARMLFNDFLSSSDPKQNNLYKGLEAFKGLK 342
QY 264 LEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGS LASVLAQYSPQL 323
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 343 LBETTYNKDVAKALLKASLDMENKTKKASLYSTNNGNMYTSSL-GCLASLSHHS AQEL 401

```


Fri Jun 25 07:33:52 2004

us-10-622-516-2.ra1

Page 8

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 198
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-401-064-198

Query Match          27.3%; Score 686; DB 4; Length 168;
Best Local Similarity 100.0%; Pred. No. 2.5e-61;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MPGLPLNAAECWPKDVGIVALEITFPSSQVVDQAELKKDGVDAKGYTTIGLGAKMGFCT 60
        |||
DB       39 MPGLPLNAAECWPKDVGIVALEITFPSSQVVDQAELKKDGVDAKGYTTIGLGAKMGFCT 98
        |||

QY      61 DREDINSLCMTVVQNLMERNNLSDYDCIGRLVEGTETIIDXSKSVKTNLMQLFEESGNTDI 120
        |||
DB       99 DREDINSLCMTVVQNLMERNNLSDYDCIGRLVEGTETIIDXSKSVKTNLMQLFEESGNTDI 158
        |||

QY      121 EGDITNACY 130
        |||
DB       159 EGDITNACY 168

RESULT 15
US-09-107-532A-5369
; Sequence 5369, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107, 532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085, 598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinfiello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 893-5007
; TELEFAX: (781) 893-8277
; INFORMATION FOR SEQ ID NO: 5369:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...388
; SEQUENCE DESCRIPTION: SEQ ID NO: 5369:

```

[illegible]

Search completed: June 24, 2004, 13:22:33
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2004, 11:14:21 ; Search time 97 Seconds
(without alignments)
1392.347 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MGSLPLNAEACWPKDVGIV.....PRLPATAEPPEAAVISNGEH 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2511	100.0	478	5	ABG32726	Abg32726 Human hyd
2	2511	100.0	478	6	ABU08379	Abu08379 Human HMG
3	2480	98.8	520	6	ABR54259	Abri54259 Human NOV
4	2480	98.8	520	6	ABR54263	Abri54263 Human NOV
5	2480	98.8	520	6	ABR54261	Abri54261 Human NOV
6	2480	98.8	520	6	ABR54270	Abri54270 Human NOV
7	2480	98.8	520	6	ABR54268	Abri54268 Human NOV
8	2480	98.8	520	6	ABR54269	Abri54269 Human NOV
9	2480	98.8	520	6	ABR54258	Abri54258 Human NOV
10	2480	98.8	520	6	ABR54267	Abri54267 Human NOV
11	2480	98.8	520	6	ABR54257	Abri54257 Human NOV
12	2480	98.8	520	6	ABR54265	Abri54265 Human NOV
13	2480	98.8	520	7	ADE61135	Ade61135 Human Pro
14	2480	98.8	520	7	ADE61139	Ade61139 Human Pro
15	2480	98.8	520	7	ADE38395	Ade38395 Human pro
16	2480	98.8	529	6	ABR54262	Abri54262 Human NOV
17	2475	98.6	519	6	ABR54260	Abri54260 Human NOV
18	2475	98.6	524	6	ABR54264	Abri54264 Human NOV
19	2475	98.6	524	6	ABR54272	Abri54272 Human NOV
20	2475	98.6	525	6	ABR54271	Abri54271 Human NOV
21	2475	98.6	525	6	ABR54266	Abri54266 Human NOV
22	2439	97.1	520	7	ADD14146	Add14146 Human src
23	2362	94.1	520	6	ABM04791	Abm04791 Rat cytos
24	2362	94.1	520	7	ADE61133	Ade61133 Rat Prote
25	2362	94.1	520	7	ADE61137	Ade61137 Rat Prote

26	2097	83.5	522	2	AAW32222	Aaw32222 Avian 3-h
27	1737	69.2	346	4	ABG20556	Abg20556 Novel hum
28	1513	60.3	508	4	AAW78869	Aam78869 Human pro
29	1513	60.3	508	6	ABR47486	Abri47486 Breast ca
30	1513	60.3	508	7	ADB75350	Abd75350 Prostate
31	1513	60.3	508	7	ADE57790	Ade57790 Human Pro
32	1513	60.3	508	7	ADE57794	Ade57794 Human Pro
33	1513	60.3	518	4	AAW79853	Aam79853 Human pro
34	1510	60.1	518	7	ADE09416	Ade09416 Novel pro
35	1494	59.5	508	7	ADE57788	Ade57788 Rat Prote
36	1494	59.5	508	7	ADE57792	Ade57792 Rat Prote
37	1337	53.2	465	4	ABR66034	Abb66034 Drosophila
38	1337	53.2	465	4	ABR60545	Abb60545 Drosophila
39	1336	53.2	583	7	ADE08499	Ade08499 Novel pro
40	1286.5	51.2	453	6	AAE38232	Aae38232 Human enz
41	1016.5	40.5	460	6	ABJ26073	Abj26073 Aspergill
42	1013	40.3	464	6	AAO27022	Aao27022 Acetyl-co
43	982	39.1	461	6	ABU62389	Abu62389 Thale cre
44	981.5	39.1	476	6	ABU62384	Abu62384 Rice hydr
45	980	39.0	461	3	AAG43333	Aag43333 Arabidops

ALIGNMENTS

RESULT 1
ID ABG32726 standard; protein; 478 AA.
XX AC ABG32726;
XX DT 22-NOV-2002 (first entry)
XX DE Human hydroxymethylglutaryl-coenzyme A synthase protein.
XX KW Human; enzyme; chromosome 5; hydroxymethylglutaryl-coenzyme A synthase;
XX KW HMG-CoA synthase; cholesterologenesis; therapeutic; diagnostic; genotype;
XX KW antibody; synthase; carcinoma.
XX OS Homo sapiens.
XX PN US6436692-B1.
XX PD 20-AUG-2002.
XX PF 29-MAR-2001; 2001US-00819993.
XX PR 29-MAR-2001; 2001US-00819993.
XX PA (APPL-) APPLERA CORP.
XX PI Gong F, Yan C, Di Francesco V, Beasley EM;
XX DR WPI; 2002-689940/74.
XX DR N-PSDB; ABS54409.
XX PT New isolated nucleic acid molecule encoding hydroxymethylglutaryl-CoA
PT synthase, useful as model for the development of human therapeutic
PT targets and for identifying therapeutic proteins.
XX PS Claim 4; Fig 2; 62pp; English.
XX CC The invention discloses an isolated nucleic acid molecule encoding a
CC human hydroxymethylglutaryl-coenzyme A (HMG-CoA) synthase which is
CC important in cholesterologenesis. The polynucleotide and polypeptide are
CC useful as models for the development of human therapeutic targets, to aid
CC in the identification of therapeutic proteins and as targets for the
CC development of human therapeutic agents that modulate the activity of the
CC polypeptide in cells and tissues. The polynucleotide is useful for
CC monitoring the effectiveness of modulating compounds on the expression or
CC activity of the enzyme gene in clinical trials and in treatments, in
CC diagnostic assays for qualitative changes in expression of enzyme nucleic
CC acid, to detect mutations in enzyme genes and gene expression products,

CC such a mRNA, for testing an individual for a genotype and as a diagnostic
CC target that can be used to tailor treatment in an individual. The
CC polypeptide is useful to raise antibodies, as a reagent in assays
CC designed to quantitatively determine levels of the protein in biological
CC fluids, as markers for tissues in which the corresponding protein is
CC preferentially expressed (e.g. carcinomas), in biological assays related
CC to members of the synthase subfamily, in drug screening assays, in
CC competition binding assays, in cell-based or cell-free systems, in
CC pharmacogenomic analysis and for treating a disorder characterised by an
CC absence of, inappropriate, or unwanted expression of the polypeptide. The
CC sequence presented is the human HMG-CoA synthase protein, the gene for
CC which is located on chromosome 5
XX
SQ Sequence 478 AA;

Query Match 100.0%; Score 2511; DB 5; Length 478;
Best Local Similarity 100.0%; Pred. No. 6.4e-237;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60
Db 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNNLSDYDCIGRLFVGTTETIIDKSKSVKTNLMOLFEEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNNLSDYDCIGRLFVGTTETIIDKSKSVKTNLMOLFEEESGNTDI 120
QY 121 EGIDTTNACYGCGTAAVFNAVNWIESSSWDGLRGTHMQAHAYDFYKPKDMLSEYPIVDGKLSI 180
Db 121 EGIDTTNACYGCGTAAVFNAVNWIESSSWDGLRGTHMQAHAYDFYKPKDMLSEYPIVDGKLSI 180
QY 181 QCYLSALDRCSYVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLND 240
Db 181 QCYLSALDRCSYVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLND 240
QY 241 FLNDQNRDKNSIYSGLEAFGVDKLEDTYFDRDVEKAFMKASSELSFQKTKASILLVSNQNG 300
Db 241 FLNDQNRDKNSIYSGLEAFGVDKLEDTYFDRDVEKAFMKASSELSFQKTKASILLVSNQNG 300
QY 301 NMYTSSVYGSILASVLAQYSPQOLAGKRIGVFSGSGLAATLYSLKVTQDATPGSALDKIT 360
Db 301 NMYTSSVYGSILASVLAQYSPQOLAGKRIGVFSGSGLAATLYSLKVTQDATPGSALDKIT 360
QY 361 ASLCDLKSRLDSRTGVAADVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH 420
Db 361 ASLCDLKSRLDSRTGVAADVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH 420
QY 421 RRTYARRPTPNDTLDGEGVGLVHSNIATEHIPSBAKKVPRLPATAAPEAAVLSNGEH 478
Db 421 RRTYARRPTPNDTLDGEGVGLVHSNIATEHIPSBAKKVPRLPATAAPEAAVLSNGEH 478

RESULT 2
ABU08379 standard; protein; 478 AA.
ID ABU08379;
XX AC ABU08379;
XX DT 05-JUN-2003 (first entry)
XX DE Human HMG-CoA synthase-like enzyme.
XX KW Human; cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase;
KW HMG-CoA synthase; immune response; drug screening assay;
KW pharmacogenomic analysis; enzyme.
XX OS Homo sapiens.
XX PN US2002173018-A1.
XX PD 21-NOV-2002.
XX PF 12-JUL-2002; 2002US-00193295.

XX 29-MAR-2001; 2001US-00819993.
PR (APPL-) APPLERA CORP.
XX Gong F, Yan C, Di Francesco V, Beasley EM;
XX WPI; 2003-352594/33.
DR N-PSDB; ABX93299, ABX93300.

PT New isolated human synthase peptide and gene encoding the enzyme, useful
PT as models for developing human therapeutic targets, aid in the
PT identification of therapeutic proteins and as therapeutic targets.

XX Claim 1; Fig 2A; 65pp; English.

XX The present invention relates to the isolation of a novel human enzyme
CC that shows a high degree of similarity to human cytoplasmic 3-hydroxy-3-
CC methylglutaryl coenzyme A (HMG-CoA) synthase. The invention also
CC discloses polynucleotide sequences encoding the novel enzyme of the
CC invention. Both the polypeptide and polynucleotide sequences are useful
CC as models for the development of human therapeutics, for identifying
CC therapeutic proteins, as targets for development of human therapeutic
CC agents, and as query sequences to perform a search against sequence data
CC bases to identify other family members of related sequences. The
CC polypeptide is useful to raise antibodies or to elicit another immune
CC response, as a reagent in assays designed to quantitatively determine
CC levels of the protein in biological fluids, as markers for tissues in
CC which the corresponding protein is preferentially expressed, in drug
CC screening assays, in cell-based or cell-free systems, to identify
CC compounds that modulate synthase activity of the protein in its natural
CC state, or an altered form that causes the specific disease or pathology
CC associated with the synthase, to screen a compound for the ability to
CC stimulate or inhibit interaction between the synthase protein and a
CC molecule that normally interacts with the synthase protein, and in
CC pharmacogenomic analysis. The polynucleotide is useful for monitoring the
CC effectiveness of modulating compounds on the expression or activity of
CC the human synthase gene in clinical trials or in a treatment regimen, in
CC diagnostic assays for qualitative changes in a human synthase nucleic
CC acid that leads to a pathology, for testing an individual for a genotype
CC that while not necessarily causing a disease, nevertheless affects the
CC treatment modality, and as antisense constructs to control human synthase
CC gene expression in cells, tissues and organisms. The present sequence
CC represents human HMG-CoA synthase-like enzyme

XX Sequence 478 AA;

Query Match 100.0%; Score 2511; DB 6; Length 478;
Best Local Similarity 100.0%; Pred. No. 6.4e-237;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60
Db 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNNLSDYDCIGRLFVGTTETIIDKSKSVKTNLMOLFEEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNNLSDYDCIGRLFVGTTETIIDKSKSVKTNLMOLFEEESGNTDI 120
QY 121 EGIDTTNACYGCGTAAVFNAVNWIESSSWDGLRGTHMQAHAYDFYKPKDMLSEYPIVDGKLSI 180
Db 121 EGIDTTNACYGCGTAAVFNAVNWIESSSWDGLRGTHMQAHAYDFYKPKDMLSEYPIVDGKLSI 180
QY 181 QCYLSALDRCSYVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLND 240
Db 181 QCYLSALDRCSYVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLND 240
QY 241 FLNDQNRDKNSIYSGLEAFGVDKLEDTYFDRDVEKAFMKASSELSFQKTKASILLVSNQNG 300
Db 241 FLNDQNRDKNSIYSGLEAFGVDKLEDTYFDRDVEKAFMKASSELSFQKTKASILLVSNQNG 300
QY 301 NMYTSSVYGSILASVLAQYSPQOLAGKRIGVFSGSGLAATLYSLKVTQDATPGSALDKIT 360
Db 301 NMYTSSVYGSILASVLAQYSPQOLAGKRIGVFSGSGLAATLYSLKVTQDATPGSALDKIT 360

Db 301 NMYTSSVYGLASVLAQYSPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKIT 360
QY 361 ASLCDLKSRLDSTRTGVADPVFAENMKLREDTHHLVNYIPQSIDSLFEGTWYLVVRDEKH 420
Db 361 ASLCDLKSRLDSTRTGVADPVFAENMKLREDTHHLVNYIPQSIDSLFEGTWYLVVRDEKH 420
QY 421 RRTYARRPTPNDDTLDEGVGLVHNSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 421 RRTYARRPTPNDDTLDEGVGLVHNSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
RESULT 3
ABR54259
ID ABR54259 standard; protein; 520 AA.
XX ABR54259;
AC ABR54259;
XX 23-JUN-2003 (first entry)
DT 23-JUN-2003 (first entry)
XX Human NOV41c protein SEQ ID NO:186.
DE Human NOV41c protein SEQ ID NO:186.
XX
KW Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cyostatic; antidiabetic; antifertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeimc;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X.
XX
OS Homo sapiens.
XX
PN WO2003023001-A2.
XX 20-MAR-2003.
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028538.
XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
PA (CURA-) CTRAGEN CORP.
XX
PI Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL,
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A,
PI Crabtree J, Dipippo VA, Edinger SR, Eilsen AJ, Ellerman K;

PI Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Paturajan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zernhusen BD, Zhong M;
XX
DR WPI; 2003-313241/30.
DR N-PSDB; ACC62328.
XX
PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS Claim 1; Page 241; 460pp; English.
XX
CC The present invention describes isolated human NOX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOX proteins given in
CC ABR54167 to ABR54276. NOX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cyostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipemic activities, and can be used in gene
CC therapy. NOX proteins are useful for treating or preventing a pathology
CC associated with a NOX protein in humans and for treating a syndrome
CC associated with the human disease. NOX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberosus sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 520 AA;
Query Match 98.8%; Score 2480; DB 6; Length 520;
Best Local Similarity 91.9%; Pred. No. 8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
QY 1 MPGSLPLNAEACWPKDVGIVALEIYFSPQYVDQAELEKKGVDAGKYYTIGLGAQKMGFCT 60
Db 1 MPGSLPLNAEACWPKDVGIVALEIYFSPQYVDQAELEKKGVDAGKYYTIGLGAQKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNNLSDYDCLGRLEVGTEITIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNNLSDYDCLGRLEVGTEITIDKSKSVKTNLMQLFEESGNTDI 120
QY 121 EGIIDTTNACYGGTAAVFNAVWIBSSWD----- 149
Db 121 EGIIDTTNACYGGTAAVFNAVWIBSSWDGRYALVVAGDIAYATGNARPTGGVGAVALL 180
QY 150 -----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKI 198
Db 181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVYKLEDTYFDRDVEKAFMKASSELFSQKTKASLIVSNONGNMYTSSVYGSGLASVLAQY 318
Db 301 FGDVYKLEDTYFDRDVEKAFMKASSELFSQKTKASLIVSNONGNMYTSSVYGSGLASVLAQY 360
QY 319 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSTRTGVAP 378

Db 361 SPQOLAGKRIGVFSYSGSLATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWLYVRVDEKRRRTYARRPTPNDDTLEDEG 438
Db 421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWLYVRVDEKRRRTYARRPTPNDDTLEDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
Db 481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
RESULT 4
ABRS4263
ID ABR54263 standard; protein; 520 AA.
XX ABR54263;
AC 23-JUN-2003 (first entry)
DT Human NOV41g protein SEQ ID NO:194.
XX
DE Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;
XX anorectic; immunosuppressive; cytostatic; antidiabetic; antifertility;
KW haemostatic; antiinflammatory; antiaesthetic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeamic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X.
XX
OS Homo sapiens.
XX WO2003023001-A2.
PN
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002MO-US028538.
XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
XX (CURA-) CURAGEN CORP.
PA
XX
PI Agee ML, Alsobrook JP, Anderson DW, Bergs C, Boldog FL,
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A,
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;

PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigara M, Patturajan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkels RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zernusen BD, Zhong M;
XX WPI: 2003-313241/30.
DR N-PSDB; ACC62332.
DR
XX
PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
XX Claim 1; Page 243; 460pp; English.
PS
XX
XX The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiaesthetic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeamic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberosus sclerosis, scleroderma, obesity, metabolic
CC congenital adrenal hyperplasia, prostate cancer, fertility, haemophilia,
CC disorders, neoplasm, lymphoma, uterus cancer, cancer, diabetes, metabolic
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 520 AA;
Query Match 98.8%; Score 2480; DB 6; Length 520;
Best Local Similarity 91.9%; Pred. No. 8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
QY 1 MPGLPLNAEACWPKDVGIVALEIYFSPQYVDAQALEKYDGDVAGKYYTIGLQAKMGFCT 60
Db 1 MPGLPLNAEACWPKDVGIVALEIYFSPQYVDAQALEKYDGDVAGKYYTIGLQAKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDSRSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDSRSVKTNLMQLFEESGNTDI 120
QY 121 EGIDTTNACYGTAAVFNAVNWIESSWDGRYALVAGDIAYVATGNARPTGCGAVALL 149
Db 121 EGIDTTNACYGTAAVFNAVNWIESSWDGRYALVAGDIAYVATGNARPTGCGAVALL 180
QY 150 -----GLRGTMOHAYDYFKPDMTSEYPIVDGKLSIQCYLSALDRCYCYCKKI 198
Db 181 IGPNAPLIFERGLRGTHMOHAYDYFKPDMTSEYPIVDGKLSIQCYLSALDRCYCYCKKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMTHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQKEGNDKDFTLNDFGFMTHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVYKLEDTYFPRDVEKAFPMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSILASVLAQY 318
Db 301 FGDVYKLEDTYFPRDVEKAFPMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSILASVLAQY 360
QY 319 SPQOLAGKRIGVFSYSGSLATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378

Db 361 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVFAENMKLREDTHHLVNIYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLVNIYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520
RESULT 5
ABR54261
ID ABR54261 standard; protein; 520 AA.
XX ABR54261;
AC ABR54261;
XX 23-JUN-2003 (first entry)
DT 23-JUN-2003 (first entry)
XX Human NOV41e protein SEQ ID NO:190.
DE Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cycostatic; antidiabetic; antifertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeimic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X.
XX Homo sapiens.
OS WO2003023001-A2.
XX 20-MAR-2003.
PD 09-SEP-2002; 2002WO-US028538.
XX 07-SEP-2001; 2001US-0318120P.
XX 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX (CURA-) CURAGEN CORP.
PA Aggee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL,
XX Burges CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A,
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;

PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyanakar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Paturajan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zernhusen BD, Zhong M;
XX WPI; 2003-313241/30.
DR N-PSDB; ACC62330.
XX Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX Claim 1; Page 242; 460pp; English.
PS The present invention describes isolated human NOVX proteins, where X is
XX 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cycostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeimic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberosus sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 520 AA;
Query Match 98.8%; Score 2480; DB 6; length 520;
Best Local Similarity 91.9%; Pred.No.8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
QY 1 MPGSLPLNAEACWPKDVGIVALEITYFPSQYVDQAELEKYDGDAGKYTIGLGQAKMGFCT 60
Db 1 MPGSLPLNAEACWPKDVGIVALEITYFPSQYVDQAELEKYDGDAGKYTIGLGQAKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTEITIDKSKSVKTNLMQLFEESSGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTEITIDKSKSVKTNLMQLFEESSGNTDI 120
QY 121 EGDITTNACYGGTAAVFNAVWIESSSWD----- 149
Db 121 EGDITTNACYGGTAAVFNAVWIESSSWDGRYALVVGDIAYATGNARPTGCVGAVALL 180
QY 150 -----GLRGTHMQAHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKI 198
Db 181 IGPNAPLIFERGLRGTHMQAHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKI 240
QY 199 HAQWQEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVXLBDITYFDRDVEKAFMKASSELFSOKTKASLVSNGNGMYTSSVYGSLASVLAQY 318
Db 301 FGDVXLBDITYFDRDVEKAFMKASSELFSOKTKASLVSNGNGMYTSSVYGSLASVLAQY 360
QY 319 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378

Db 361 SPQOLAKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRITYARRPTPNDDTILDEG 438
Db 421 DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRITYARRPTPNDDTILDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPPEAAVINGEH 478
Db 481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPPEAAVINGEH 520
RESULT 6
ABR54270
ID ABR54270 standard; protein; 520 AA.
XX ABR54270;
AC ABR54270;
XX 23-JUN-2003 (first entry)
XX Human NOV4ln protein SEQ ID NO:208.
DE Human NOV4ln protein SEQ ID NO:208.
XX Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cyostatic; antidiabetic; antifertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X.
XX Homo sapiens.
XX WO2003023001-A2.
XX 20-MAR-2003.
PD 09-SEP-2002; 2002WO-US028538.
PF 07-SEP-2001; 2001US-0318120P.
XX 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX (CURA-) CURAGEN CORP.
PA
XX Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL,
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A,
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;

PI Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Kramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Paturajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zethusen BD, Zhong M;
XX WPI; 2003-313241/30.
DR N-PSDB; ACC62339.
DR
XX Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX Claim 1; Page 247; 460pp; English.
PS
XX The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cyostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipemic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberosus sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
XX Sequence 520 AA;
SQ
Query Match 98.8%; Score 2480; DB 6; length 520;
Best Local Similarity 91.9%; Pred. No. 8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
QY 1 M P G S L P L N A E A C W P K D V G I A L E I Y F P S Q Y V D Q A L E K Y D G V D A G K Y T I G L G A K M G F C T 60
Db 1 M P G S L P L N A E A C W P K D V G I A L E I Y F P S Q Y V D Q A L E K Y D G V D A G K Y T I G L G A K M G F C T 60
QY 61 D R E D I N S L C M T V V Q N L M E R N N L S Y D C I G R L E V G T E T I I D K S V K T N L M Q L F E S G N T D I 120
Db 61 D R E D I N S L C M T V V Q N L M E R N N L S Y D C I G R L E V G T E T I I D K S V K T N L M Q L F E S G N T D I 120
QY 121 E G I D T N A C Y G T A A V F N A V N W I E S S S W D G R Y A L V V A G D I A V Y A T G N A R P T G V G A V A L L 180
Db 121 E G I D T N A C Y G T A A V F N A V N W I E S S S W D G R Y A L V V A G D I A V Y A T G N A R P T G V G A V A L L 180
QY 150 -----G I R G T H M Q H A Y D F Y K P D M L S E Y P I V D G K L S I O C Y L S A L D R C Y S Y C K I 198
Db 181 I G P N A P L I F E R G I R G T H M Q H A Y D F Y K P D M L S E Y P I V D G K L S I O C Y L S A L D R C Y S Y C K I 240
QY 199 H A O W O K E G N D K D F T L N D F G F M I F H S P Y C K L V O K S L A R M L L N D F L N D Q N R D K N S I Y S G L E A 258
Db 241 H A O W O K E G N D K D F T L N D F G F M I F H S P Y C K L V O K S L A R M L L N D F L N D Q N R D K N S I Y S G L E A 300
QY 259 F G D V K L E D T Y F D R D V E K A F M K A S S E L F S O K T R A S L L V S N O N G N M Y T S S V Y G S L A V L A O Y 318
Db 301 F G D V K L E D T Y F D R D V E K A F M K A S S E L F S O K T R A S L L V S N O N G N M Y T S S V Y G S L A V L A O Y 360
QY 319 S P O Q L A K R I G V F S Y S G S L A A T L Y S L K V T O D A T P G S A L D K I T A S L C D L K S R L D S R T G V A P 378

Db 361 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSRTGVAP 420
QY 379 DYFAENMKLREDTHHLVNYIPQSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDEG 438
Db 421 DYFAENMKLREDTHHLVNYIPQSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520
RESULT 7
ABR54268
ID ABR54268 standard; protein; 520 AA.
XX ABR54268;
AC
XX
DT 23-JUN-2003 (first entry)
XX
DE Human NOV411 protein SEQ ID NO:204.
XX
KW Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cyostatic; antidiabetic; antifertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X.
XX
OS Homo sapiens.
XX
PN WO2003023001-A2.
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028538.
XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Agee ML, Alsbrook JP, Anderson DW, Berghs C, Boldog FL,
PI Burgess CE, Caeman SJ, Catterton E, Chant JS, Chaudhuri A,
PI Crabtree J, Dapipo VA, Edinger SR, Eisen AJ, Ellerman K;

PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malysankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patutarajan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zernusen BD, Zhong M;
XX
DR WPI; 2003-313241/30.
DR N-PSDB; ACC62337.
XX
PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS Claim 1; Page 246; 460pp; English.
XX
CC The present invention describes isolated human NOX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOX proteins given in
CC ABR54167 to ABR54276. NOX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cyostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeic activities, and can be used in gene
CC therapy. NOX proteins are useful for treating or preventing a pathology
CC associated with a NOX protein in humans and for treating a syndrome
CC associated with the human disease. NOX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberosus sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 520 AA;
Query Match 98.8%; Score 2480; DB 6; length 520;
Best Local Similarity 91.9%; Pred No. 8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
QY 1 MPGLPLNAEACWPKDVGIVALEIYFPSPQYVDQAELEKYDGDAGKYTIGLGOAKMGFCT 60
Db 1 MPGLPLNAEACWPKDVGIVALEIYFPSPQYVDQAELEKYDGDAGKYTIGLGOAKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTEITIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNLSYDCIGRLEVGTEITIDKSKSVKTNLMQLFEESGNTDI 120
QY 121 EGIPTTNACYCGTAAVFNAVNMTIESSWD----- 149
Db 121 EGIPTTNACYCGTAAVFNAVNMTIESSWDGRYALVVAAGDIAVYATGNARPTGGVAVALL 180
QY 150 -----GLRGTHMOHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKI 198
Db 181 IGPNAPLIFERGLRGTHMOHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKI 240
QY 199 HAQWQEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVRLDPTYFDRDVEKAFMKASSELFSOKTKASLVSNOGNMYTSSVYGSILASVLAQY 318
Db 301 FGDVRLDPTYFDRDVEKAFMKASSELFSOKTKASLVSNOGNMYTSSVYGSILASVLAQY 360
QY 319 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSRTGVAP 378

Db 361 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVFAENMKLRDTHHLVNYIPQGSIDSLFEGTWYLVRYDEKRRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLRDTHHLVNYIPQGSIDSLFEGTWYLVRYDEKRRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIPSAPKKVRLPATAPPEAAVISINGEH 478
Db 481 VGLVHSNIATEHIPSAPKKVRLPATAPPEAAVISINGEH 520

RESULT 8
ABR54269
ID ABR54269 standard; protein; 520 AA.
XX
AC ABR54269;

DT 23-JUN-2003 (first entry)
XX
DE Human NOV41m protein SEQ ID NO:206.

XX Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cyostatic; antidiabetic; antiinfertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberos sclerosi; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X.

XX OS Homo sapiens.
XX PN WO2003023001-A2.
XX PD 20-MAR-2003.
XX

PF 09-SEP-2002; 2002WO-US028538.

XX 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.

XX PA (CURA-) CURAGEN CORP.

XX Aggee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL,
PI Burgess CE, Casman SJ, Catcerton E, Chant JS, Chaudhuri A,
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;

PI Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W,
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malayankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Patturajan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twonlow N, Vernet CAM, Voss EZ;
PI Zernusen BD, Zhong M;
XX
DR WPI, 2003-313241/30.
DR N-PSDB; ACC62338.

Novel human proteins and nucleic acid encoding the proteins, useful for
diagnosis, treatment and prevention of disorders involving the human
protein or nucleic acid e.g. cardiac and neurological disorders.

Claim 1; Page 246; 460pp; English.

XX The present invention describes isolated human NOX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOX proteins given in
CC ABR54167 to ABR54276. NOX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cyostatic,
CC antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaemic activities, and can be used in gene
CC therapy. NOX proteins are useful for treating or preventing a pathology
CC associated with a NOX protein in humans and for treating a syndrome
CC associated with the human disease. NOX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberos sclerosi, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX

SQ Sequence 520 AA;

Query Match 98.8%; Score 2480; DB 6; Length 520;
Best Local Similarity 91.9%; Pred. No. 8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MPGLPLNBAACWPKDVGIVALEIYFPQYVDQALEKYDGDAGKYTTIGLQAKMGFCT 60
Db 1 MPGLPLNBAACWPKDVGIVALEIYFPQYVDQALEKYDGDAGKYTTIGLQAKMGFCT 60
QY 61 DREDINSLCMTVQVONLMERNLSYDCIGRLVGTETIIDRSKSVKTNLMQLFEESENTDI 120
Db 61 DREDINSLCMTVQVONLMERNLSYDCIGRLVGTETIIDRSKSVKTNLMQLFEESENTDI 120
QY 121 EGIDTTNACYGTAAVFNAVNWIESSWD----- 149
Db 121 EGIDTTNACYGTAAVFNAVNWIESSWDGRYALVAGDIAVYATGNARPTGGVAVALL 180
QY 150 -----GLRGTHQHAHYDFYKPKMLSEYPIVDGKLISIQCYLSALDRCYSVYCKKI 198
Db 181 IGPNAPLIFERGLRGTHQHAHYDFYKPKMLSEYPIVDGKLISIQCYLSALDRCYSVYCKKI 240
QY 199 HAQWQKEGNDKDFTLNDRGFMIHFSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQKEGNDKDFTLNDRGFMIHFSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLVSNQNGNMTSSVYGSILASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLVSNQNGNMTSSVYGSILASVLAQY 360
QY 319 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378

```

Db      361 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSRTGVAP 420
QY      379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDG 438
        |||||
Db      421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYARRPTPNDDTLDG 480
QY      439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
        |||||
Db      481 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520

RESULT 9
ABR54258
ID      ABR54258 standard; protein; 520 AA.
XX
AC      ABR54258;
XX
DT      23-JUN-2003 (first entry)
XX
DE      Human NOVA1b protein SEQ ID NO:184.
XX
KW      Human; NOVA; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW      anorectic; immunosuppressive; cyostatic; antidiabetic; antifertility;
KW      haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW      neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeamic;
KW      gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW      congenital heart defect; aortic stenosis; valve disease; transplantation;
KW      tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW      prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW      fertility; haemophilia; hypercoagulation; graft versus host disease;
KW      idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW      Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW      cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW      immune disorder; haematopoietic disorder; dyslipidaemia;
        metabolic syndrome X.
XX
OS      Homo sapiens.
XX
PN      WO2003023001-A2.
XX
PD      20-MAR-2003.
XX
PF      09-SEP-2002; 2002WO-US028538.
XX
PR      07-SEP-2001; 2001US-0318120P.
PR      07-SEP-2001; 2001US-0318184P.
PR      10-SEP-2001; 2001US-0318430P.
PR      17-SEP-2001; 2001US-0322636P.
PR      17-SEP-2001; 2001US-0322781P.
PR      17-SEP-2001; 2001US-0322816P.
PR      17-SEP-2001; 2001US-0322817P.
PR      19-SEP-2001; 2001US-0323519P.
PR      20-SEP-2001; 2001US-0323631P.
PR      20-SEP-2001; 2001US-0323636P.
PR      25-SEP-2001; 2001US-0324969P.
PR      25-SEP-2001; 2001US-0325091P.
PR      26-SEP-2001; 2001US-0324990P.
PR      14-DEC-2001; 2001US-0341144P.
PR      26-FEB-2002; 2002US-0359599P.
PR      05-MAR-2002; 2002US-0361663P.
PR      03-MAY-2002; 2002US-0377908P.
PR      17-MAY-2002; 2002US-0381483P.
PR      29-MAY-2002; 2002US-0383863P.
PR      02-JUL-2002; 2002US-0393332P.
PR      17-JUL-2002; 2002US-0396412P.
PR      13-AUG-2002; 2002US-0403517P.
PR      06-SEP-2002; 2002US-00236417.
XX
PA      (CURA-) CURAGEN CORP.
XX
PI      Agee ML, Albrock JP, Anderson DW, Berghs C, Boldog FL,
PI      Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A,
PI      Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
```

```

PI      Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI      Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI      Malyanakar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Paturejan M;
PI      Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI      Spaderna SK, Spytek KA, Taupier RJ, Twonlow N, Vernet CAM, Voss EZ;
PI      Zernhusen BD, Zhong M;
XX
DR      WPI; 2003-313241/30.
DR      N-PSDB; ACC62327.
XX
PT      Novel human proteins and nucleic acid encoding the proteins, useful for
PT      diagnosis, treatment and prevention of disorders involving the human
PT      protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS      Claim 1; Page 241; 460pp; English.
XX
CC      The present invention describes isolated human NOVX proteins, where X is
CC      1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC      ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC      hypotensive, dermatological, anorectic, immunosuppressive, cyostatic,
CC      antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC      antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC      antiparkinsonian and antilipaeamic activities, and can be used in gene
CC      therapy. NOVX proteins are useful for treating or preventing a pathology
CC      associated with a NOVX protein in humans and for treating a syndrome
CC      associated with the human disease. NOVX nucleic acids, proteins and
CC      antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC      atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC      valve disease, tuberosus sclerosis, scleroderma, obesity, transplantation,
CC      congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC      disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC      hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC      disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC      infectious disease, anorexia, cancer-associated cachexia, cancer,
CC      Alzheimer's disease, Parkinson's disease, immune disorders,
CC      haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC      ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC      sequences, which are used in examples from the present invention.
CC      ABR54277 represents a human trypsinogen protein given in comparison with
CC      the human NOV35b protein in the exemplification of the present invention
XX
SQ      Sequence 520 AA;

Query Match          98.8%; Score 2480; DB 6; Length 520;
Best Local Similarity 91.9%; Pred. No. 8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGAQKMGFCT 60
        |||||
Db      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGAQKMGFCT 60
QY      61 DREDINSLCMTVVQNLMERNLSDYDCIGRLVGTETITIDSKSVKTNLMQLFEESGNTDI 120
        |||||
Db      61 DREDINSLCMTVVQNLMERNLSDYDCIGRLVGTETITIDSKSVKTNLMQLFEESGNTDI 120
QY      121 EGDITTNACYGGTAAPFNAVNWIESSSWD----- 149
        |||||
Db      121 EGDITTNACYGGTAAPFNAVNWIESSSWDGRYALVAVAGDIAYATGNARPTGCVAVALL 180
QY      150 -----GLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
        |||||
Db      181 IGPNAPLIFERGLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240
QY      199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGIEA 258
        |||||
Db      241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGIEA 300
QY      259 FGDVPLEDTYFDRDVEKAFMKASSELFSQKTASLIVSNONGNMYTSSVYGSLASVLAQY 318
        |||||
Db      301 FGDVPLEDTYFDRDVEKAFMKASSELFSQKTASLIVSNONGNMYTSSVYGSLASVLAQY 360
QY      319 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSRTGVAP 378
        |||||
```


Db 361 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVEAENMKLRDPTHTHLVNYIPQGSIDSLFEGTWTLVVRVDEKHKRTYARPTPNDDTLDEG 438
Db 421 DVEAENMKLRDPTHTHLVNYIPQGSIDSLFEGTWTLVVRVDEKHKRTYARPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHTIPSPAKKVPRLPATAAEPAAVINSNGEH 478
Db 481 VGLVHSNIATEHTIPSPAKKVPRLPATAAEPAAVINSNGEH 520
RESULT 10
ABR54267
ID ABR54267 standard; protein; 520 AA.
XX
AC ABR54267;
XX
XX 23-JUN-2003 (first entry)
XX
DE Human NOV41k protein SEQ ID NO:202.
XX
KW Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cytostatic; antidiabetic; antinfertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantacion;
KW tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X.
XX
XX Homo sapiens.
XX
XX WO2003023001-A2.
XX
XX 20-MAR-2003.
XX
XX 09-SEP-2002; 2002WO-US028538.
XX
XX 07-SEP-2001; 2001US-0318120P.
XX 07-SEP-2001; 2001US-0318184P.
XX 10-SEP-2001; 2001US-0318430P.
XX 17-SEP-2001; 2001US-0322636P.
XX 17-SEP-2001; 2001US-0322781P.
XX 17-SEP-2001; 2001US-0322816P.
XX 17-SEP-2001; 2001US-0322817P.
XX 19-SEP-2001; 2001US-0323519P.
XX 20-SEP-2001; 2001US-0323631P.
XX 20-SEP-2001; 2001US-0323636P.
XX 25-SEP-2001; 2001US-0324969P.
XX 25-SEP-2001; 2001US-0325091P.
XX 26-SEP-2001; 2001US-0324990P.
XX 14-DEC-2001; 2001US-0341144P.
XX 26-FEB-2002; 2002US-0359599P.
XX 05-MAR-2002; 2002US-0361663P.
XX 03-MAY-2002; 2002US-0377908P.
XX 17-MAY-2002; 2002US-0381483P.
XX 29-MAY-2002; 2002US-0383863P.
XX 02-JUL-2002; 2002US-0393332P.
XX 17-JUL-2002; 2002US-0396412P.
XX 13-AUG-2002; 2002US-0403517P.
XX 06-SEP-2002; 2002US-00236417.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Aggee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL,
XX Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A,
XX Crabtree J, Dippio VA, Edinger SR, Eisen AJ, Ellerman K,

PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Patturajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zernusen BD, Zhong M;
XX
XX WPI; 2003-313241/30.
XX
XX N-PSDB; ACC62336.
XX
XX Novel human proteins and nucleic acid encoding the proteins, useful for
XX diagnosis, treatment and prevention of disorders involving the human
XX protein or nucleic acid e.g. cardiac and neurological disorders.
XX
XX Claim 1; Page 245; 460pp; English.
XX
XX The present invention describes isolated human NOVX proteins, where X is
XX 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
XX ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
XX hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
XX antidiabetic, antinfertility, haemostatic, antiinflammatory, anti-HIV,
XX antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
XX antiparkinsonian and antilipemic activities, and can be used in gene
XX therapy. NOVX proteins are useful for treating or preventing a pathology
XX associated with a NOVX protein in humans and for treating a syndrome
XX associated with the human disease. NOVX nucleic acids, proteins and
XX antibodies can be used in the treatment and diagnosis of cardiomyopathy,
XX atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
XX valve disease, tuberosus sclerosis, scleroderma, obesity, transplantacion,
XX congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
XX disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
XX hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
XX disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
XX infectious disease, anorexia, cancer-associated cachexia, cancer,
XX Alzheimer's disease, Parkinson's disease, immune disorders,
XX haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
XX ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
XX sequences, which are used in examples from the present invention.
XX ABR54277 represents a human trypsinogen protein given in comparison with
XX the human NOV35b protein in the exemplification of the present invention
XX
XX SQ Sequence 520 AA;
XX
XX Query Match 98.8%; Score 2480; DB 6; Length 520;
XX Best Local Similarity 91.9%; Pred. No. 8.1e-234;
XX Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
QY 1 MPGSLPLNAEACWPKDVGIVALEIFPSQYVDQAELEKXGVDAGKYTTIGLGAQKMGFCT 60
Db 1 MPGSLPLNAEACWPKDVGIVALEIFPSQYVDQAELEKXGVDAGKYTTIGLGAQKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNNLSDYDCIGRLVGTETITIDSKSVKTNLMQLFEESENTDI 120
Db 61 DREDINSLCMTVVQNLMERNNLSDYDCIGRLVGTETITIDSKSVKTNLMQLFEESENTDI 120
QY 121 EGIDTTNACYGGTAAVFNAVNMIESSWD----- 149
Db 121 EGIDTTNACYGGTAAVFNAVNMIESSWDGRYALVAGDIAVYATGNARPTGCVAVALL 180
QY 150 -----GLRGTHNQHAYDFYKPDMLSEYPYVDGKLISIQCYLSALDRCYSVYCKKI 198
Db 181 IGPNAPLIFERGLRGTHNQHAYDFYKPDMLSEYPYVDGKLISIQCYLSALDRCYSVYCKKI 240
QY 199 HAQWQKEGNDKDFTLNDRGFMIHFSPYCKLVQKSLARMLNDFLNDQNRKNSIYSGLEA 258
Db 241 HAQWQKEGNDKDFTLNDRGFMIHFSPYCKLVQKSLARMLNDFLNDQNRKNSIYSGLEA 300
QY 259 FGDVKLEBDTYFDRDVEKAFMKASSELFQSOKTKASILVSNONGNMYTSSVYGSLASVLAQY 318
Db 301 FGDVKLEBDTYFDRDVEKAFMKASSELFQSOKTKASILVSNONGNMYTSSVYGSLASVLAQY 360
QY 319 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378

Db	361	SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSTRGVAP	420
Qy	379	DVFAENMKLRDETHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG	438
Db	421	DVFAENMKLRDETHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG	480
Qy	439	VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH	478
Db	481	VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH	520
RESULT 11			
ABR54257			
ID	ABR54257	standard; protein; 520 AA.	
XX	AC	ABR54257;	
XX	DT	23-JUN-2003 (first entry)	
XX	DE	Human NOV41a protein SEQ ID NO:182.	
XX	KW	Human; NOX; antiatherosclerotic; hypotensive; cardiast; dermatological;	
KW		anorectic; immunosuppressive; cytostatic; antidiabetic; antifertility;	
KW		haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;	
KW		neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeic;	
KW		gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;	
KW		congenital heart defect; aortic stenosis; valve disease; transplantation;	
KW		tuberos sclerosis; obesity; congenital adrenal hyperplasia; diabetes;	
KW		prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;	
KW		fertility; haemophilia; hypercoagulation; graft versus host disease;	
KW		idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;	
KW		Crohn's disease; multiple sclerosis; infectious disease; cancer;	
KW		cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;	
KW		immune disorder; haematopoietic disorder; dyslipidaemia;	
metabolic syndrome X.			
XX	OS	Homo sapiens.	
XX	PN	WO2003023001-A2.	
XX	PD	20-MAR-2003.	
XX	PF	09-SEP-2002; 2002WO-US028538.	
XX	PR	07-SEP-2001; 2001US-0318120P.	
PR		07-SEP-2001; 2001US-0318184P.	
PR		10-SEP-2001; 2001US-0318430P.	
PR		17-SEP-2001; 2001US-0322636P.	
PR		17-SEP-2001; 2001US-0322781P.	
PR		17-SEP-2001; 2001US-0322816P.	
PR		17-SEP-2001; 2001US-0322817P.	
PR		19-SEP-2001; 2001US-0323519P.	
PR		20-SEP-2001; 2001US-0323631P.	
PR		20-SEP-2001; 2001US-0323636P.	
PR		25-SEP-2001; 2001US-0324969P.	
PR		25-SEP-2001; 2001US-0325091P.	
PR		26-SEP-2001; 2001US-0324990P.	
PR		14-DEC-2001; 2001US-0341144P.	
PR		26-FEB-2002; 2002US-0359599P.	
PR		05-MAR-2002; 2002US-0361663P.	
PR		03-MAY-2002; 2002US-0377908P.	
PR		17-MAY-2002; 2002US-0381483P.	
PR		29-MAY-2002; 2002US-0383863P.	
PR		02-JUL-2002; 2002US-0393332P.	
PR		17-JUL-2002; 2002US-0396412P.	
PR		13-AUG-2002; 2002US-0403517P.	
PR		06-SEP-2002; 2002US-00236417.	
XX	PA	(CURA-) CURAGEN CORP.	
XX	PI	Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;	
PI		Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;	
PI		Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;	

PI	Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;	
PI	Kekuda R, Khrantsov NV, Leach MD, Lopley DM, Li L, Liu X;	
PI	Malayankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Paturajan M;	
PI	Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;	
PI	Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;	
PI	Zerhusen BD, Zhong M;	
XX		
DR	WPI; 2003-313241/30.	
DR	N-PSDB; ACC62326.	
XX		
PT	Novel human proteins and nucleic acid encoding the proteins, useful for	
PT	diagnosis, treatment and prevention of disorders involving the human	
PT	protein or nucleic acid e.g. cardiac and neurological disorders.	
XX		
PS	Claim 1; Page 240; 460pp; English.	
XX		
CC	The present invention describes isolated human NOVX proteins, where X is	
CC	1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in	
CC	ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiast,	
CC	hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,	
CC	antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,	
CC	antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,	
CC	antiparkinsonian and antilipemic activities, and can be used in gene	
CC	therapy. NOVX proteins are useful for treating or preventing a pathology	
CC	associated with a NOVX protein in humans and for treating a syndrome	
CC	associated with the human disease. NOVX nucleic acids, proteins and	
CC	antibodies can be used in the treatment and diagnosis of cardiomyopathy,	
CC	atherosclerosis, hypertension, congenital heart defects, aortic stenosis,	
CC	valve disease, tuberos sclerosis, scleroderma, obesity, transplantation,	
CC	congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic	
CC	disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,	
CC	hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host	
CC	disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,	
CC	infectious disease, anorexia, cancer-associated cachexia, cancer,	
CC	Alzheimer's disease, Parkinson's disease, immune disorders,	
CC	haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.	
CC	ACC62346 to ACC62465 represent PCR primers and probes for human NOVX	
CC	sequences, which are used in examples from the present invention.	
CC	ABR54277 represents a human trypsinogen protein given in comparison with	
CC	the human NOV35b protein in the exemplification of the present invention	
XX		
SQ	Sequence 520 AA;	
Query Match		
Best Local Similarity 98.8%; Score 2480; DB 6; length 520;		
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;		
Qy	1	MPGSLPLNAEACWPKDVGIVALEIFYPSQYVDQAELEKYGVDAGKYTIGLGOAKMGFCT 60
Db	1	MPGSLPLNAEACWPKDVGIVALEIFYPSQYVDQAELEKYGVDAGKYTIGLGOAKMGFCT 60
Qy	61	DREDINSLCMTVVQNLMERNNLSYDCIGRLVEGTETITIDSKSVKTNLMQLFEESGNTDI 120
Db	61	DREDINSLCMTVVQNLMERNNLSYDCIGRLVEGTETITIDSKSVKTNLMQLFEESGNTDI 120
Qy	121	EGIDTTNACYGGTAAVFNAAVWIESSWD----- 149
Db	121	EGIDTTNACYGGTAAVFNAAVWIESSWDGRRYALVAGDIAVYATGNARPTGGVGAVALL 180
Qy	150	-----GLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCSYVYCKKI 198
Db	181	IGPNAPLIFERGLRGTHMQHAYDYFKPDMLSEYPIVDGKLSIQCYLSALDRCSYVYCKKI 240
Qy	199	HAQWQEGNDKDFTLNDFGFMI FHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db	241	HAQWQEGNDKDFTLNDFGFMI FHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
Qy	259	FGDVKLEDTYFDRDVEKA FMRASSELFSOKTKASLVSNGNMYTSSVYGSIASVLAQY 318
Db	301	FGDVKLEDTYFDRDVEKA FMRASSELFSOKTKASLVSNGNMYTSSVYGSIASVLAQY 360
Qy	319	SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCLDKSRLDSTRGVAP 378

Db 361 SPQLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420
Qy 379 DVFAENMKLRDETHLVNYTPQGSIDSLFEGTWYLVVRVVEKRRRTYARRPTPNDTLDG 438
Db 421 DVFAENMKLRDETHLVNYTPQGSIDSLFEGTWYLVVRVDEKRRRTYARRPTPNDTLDG 480
Qy 439 VGLVHSNIAATEHIPSAPKVPRLPATAPPEAAVISNGEH 478
Db 481 VGLVHSNIAATEHIPSAPKVPRLPATAPPEAAVISNGEH 520

RESULT 12
ABR54265
ID ABR54265 standard; protein, 520 AA.
XX
AC ABR54265;
XX
DT 23-JUN-2003 (first entry)
XX
DE Human NOV41: protein SEQ ID NO:198.
XX

XX Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cytostatic; antidiabetic; antifertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeamic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantion;
KW tuberos sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X.
XX

OS Homo sapiens.
XX
PN WO2003023001-A2.
XX
PD 20-MAR-2003.
XX

PF 09-SEP-2002; 2002WO-US028538.
XX

PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 03-MAR-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX

PA (CURA-) CURAGEN CORP.
XX

PI Agee ML, Alsobrook JP, Anderson DW, Berghs C, Bolog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;

PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Pattnajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twonlow N, Vernet CAM, Voss EZ;
PI Zehusen BD, Zhong M;
XX
DR WPI; 2003-313241/30.
DR N-PSDB; ACC62334.
XX

PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX

PS Claim 1; Page 244; 460pp; English.
XX

CC The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeamic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberos sclerosis, scleroderma, obesity, transplantion,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX

SQ Sequence 520 AA;
XX

Query Match 98.8%; Score 2480; DB 6; Length 520;
Best Local Similarity 91.9%; Pred. No. 8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

Qy 1 MPGLPLNAEACWPKDVGIVALEIYFPSSQYVDQAELEKYDGVDAKTYTIGLGQAKGFCF 60
Db 1 MPGLPLNAEACWPKDVGIVALEIYFPSSQYVDQAELEKYDGVDAKTYTIGLGQAKGFCF 60
Qy 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEEGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEEGNTDI 120
Qy 121 EGIDTTNACYGTAAVFNAVNWTESSSWD----- 149
Db 121 EGIDTTNACYGTAAVFNAVNWTESSSWDGRYALVVAAGDIAYVATGNARPTGGVCAVALL 180
Qy 150 -----GLRGTHMQHAYDFYKPKDMLSEYPIVDGKLSIOCYLSALDRCTSVYCKKI 198
Db 181 IGENAPLIFERGLRGTHMQHAYDFYKPKDMLSEYPIVDGKLSIOCYLSALDRCTSVYCKKI 240
Qy 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
Qy 259 FGDVYKLEDYTRDRDVEKAFMKASSELFSQKTKASLIVSNQNGMYTSSVYGSILASVLAQY 318
Db 301 FGDVYKLEDYTRDRDVEKAFMKASSELFSQKTKASLIVSNQNGMYTSSVYGSILASVLAQY 360
Qy 319 SPQLAGKRIGVFSYSGSLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378

Db 361 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 438
Cc ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cc 421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520

RESULT 13
ADE61135
ID ADE61135 standard; protein; 520 AA.
XX
AC ADE61135;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human Protein Q01581, SEQ ID NO 7053.
XX
KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; Q01581.
XX
DR
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
Cc The invention discloses a composition comprising two or more isolated rat
Cc or human polynucleotides or a polynucleotide which represents a fragment,
Cc derivative or allelic variation of the nucleic acid sequence. Also
Cc claimed are a vector comprising the novel polynucleotide, a host cell
Cc comprising the vector, a method for identifying a nucleotide sequence
Cc which is differentially regulated in an animal subjected to pain and a
Cc kit to perform the method, an array, a method for identifying an agent
Cc that increases or decreases the expression of the polynucleotide sequence
Cc that is differentially expressed in neuronal tissue of a first animal
Cc subjected to pain, a method for identifying a compound which regulates
Cc the expression of a polynucleotide sequence which is differentially
Cc expressed in an animal subjected to pain, a method for identifying a
Cc compound that regulates the activity of one or more of the
Cc polynucleotides, a method for producing a pharmaceutical composition, a
Cc method for identifying a compound or small molecule that regulates the
Cc activity in an animal of one or more of the polypeptides given in the
Cc specification, a method for identifying a compound useful in treating
Cc pain and a pharmaceutical composition comprising the one or more
Cc polypeptides or their antibodies. The polynucleotide or the compound that
Cc modulates its activity is useful for preparing a medicament for treating
Cc pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
Cc injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
Cc therapy). The sequence presented is a human protein (shown in Table 2 of

Cc the specification) which is differentially expressed during pain. Note:
Cc The sequence data for this patent did not form part of the printed
Cc specification, but was obtained in electronic form directly from WIPO at
Cc ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 520 AA;

Query Match 98.8%; Score 2480; DB 7; Length 520;
Best Local Similarity 91.9%; Pred. No. 8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKXDVADGKYTIGLGAIXGFCF 60
Db 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKXDVADGKYTIGLGAIXGFCF 60
QY 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLVEGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLVEGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY 121 EGDITTNACYGCTAAVFNANWMISSSWD----- 149
Db 121 EGDITTNACYGCTAAVFNANWMISSSWDGRYALVAVAGDIAVATGNARPTGGVAVALL 180
QY 150 -----GLRGTHQHAAYDFYKPDMLSEYPIVDGSLIQCYLSALDRCYSVCKKI 198
Db 181 IGPNAPLIFERGLRGTHQHAAYDFYKPDMLSEYPIVDGSLIQCYLSALDRCYSVCKKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMNLNDFLNDONRDKNSIYSGIEA 258
Db 241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMNLNDFLNDONRDKNSIYSGIEA 300
QY 259 FGDVKLEDTYFDRDVEKAFMKASSELSQKTKASLIVSNONGNMYTSSVYGSASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASSELSQKTKASLIVSNONGNMYTSSVYGSASVLAQY 360
QY 319 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db 361 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 520

RESULT 14
ADE61139
ID ADE61139 standard; protein; 520 AA.
XX
AC ADE61139;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human Protein Q01581, SEQ ID NO 7057.
XX
KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.

XX (GENO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
PI Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; Q01581.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 520 AA;

Query Match 98.8%; Score 2480; DB 7; Length 520;
Best Local Similarity 91.9%; Pred. No. 8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MPGSLPINABACWPKDVGIVALEIYFPPSQYVDQAELEKYPGDVDAKYYTIGLQAKMGFCT 60
Db 1 MPGSLPINABACWPKDVGIVALEIYFPPSQYVDQAELEKYPGDVDAKYYTIGLQAKMGFCT 60

QY 61 DREDINSLCMTVVQNLMEERNLSYDCIGRIEVTGTEITIDKSKSVKTNLMQIFEESENTDI 120
Db 61 DREDINSLCMTVVQNLMEERNLSYDCIGRIEVTGTEITIDKSKSVKTNLMQIFEESENTDI 120

QY 121 EGIDTTNACYGCTAAVFNAVNWIESSSWD----- 149
Db 121 EGIDTTNACYGCTAAVFNAVNWIESSSWDGRYALVVGDIAYATGNARPTGCVGAVALL 180

QY 150 -----GLRGTHMQAHAYDFYKPKDMLSEYPIVDGKLSIQCYLSALDRCTSVYCKKI 198
Db 181 IGPNAPLIFERGLRGTHMQAHAYDFYKPKDMLSEYPIVDGKLSIQCYLSALDRCTSVYCKKI 240

QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300

QY 259 FGDVKLEDTYFDRDVEKAFMKASSELFSOKTKASLLVSNQNGMWYTSSVYGSLASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASSELFSOKTKASLLVSNQNGMWYTSSVYGSLASVLAQY 360

QY 319 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378

Db 361 SPQQLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

QY 379 DVFAENMKLRBDTHLVVNIYIPQGSIDSLFEGTWYLVRVDEKGRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLRBDTHLVVNIYIPQGSIDSLFEGTWYLVRVDEKGRRTYARRPTPNDDTLDEG 480

QY 439 VGLVHSNIAATEHIPSPAKKVPRLPATAAPEPAVISNGEH 478
Db 481 VGLVHSNIAATEHIPSPAKKVPRLPATAAPEPAVISNGEH 520

RESULT 15
ADE38395
ID ADE38395 standard; protein; 520 AA.
XX
AC ADE38395;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human protein 9389 amino acid sequence.
XX
KW tumorigenic disorder; angiogenic disorder; aberrant gene expression;
KW aberrant protein activity; cytostatic; antithyroid; antidiabetic;
KW ophthalmological; cancer; breast cancer; colon cancer; lung cancer;
KW prostatic cancer; Grave's disease; diabetic retinopathy; protein 9389.
XX
OS Homo sapiens.
XX
PN WO2003065006-A2.
XX
PD 07-AUG-2003.
XX
PF 30-JAN-2003; 2003WO-US002588.
XX
PF 31-JAN-2002; 2002US-0353600P.
PR 15-MAR-2002; 2002US-0364517P.
PR 09-APR-2002; 2002US-0371075P.
PR 10-APR-2002; 2002US-0371507P.
PR 16-APR-2002; 2002US-0372984P.
PR 19-APR-2002; 2002US-0374194P.
PR 24-MAY-2002; 2002US-0382995P.
PR 31-MAY-2002; 2002US-0385023P.
PR 14-JUN-2002; 2002US-0388853P.
PR 17-JUN-2002; 2002US-0389395P.
PR 25-JUN-2002; 2002US-0391324P.
PR 15-JUL-2002; 2002US-0395944P.
PR 22-JUL-2002; 2002US-0397726P.
PR 13-AUG-2002; 2002US-0403046P.
PR 22-AUG-2002; 2002US-0405155P.
PR 27-AUG-2002; 2002US-0406361P.
PR 25-OCT-2002; 2002US-0421195P.
PR 12-NOV-2002; 2002US-0425456P.
PR 19-NOV-2002; 2002US-0427626P.
PR 10-DEC-2002; 2002US-0432122P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Hunter JI, Macbeth KJ, Tsai F, Lesoon A, Lightcap ES;
PI Williamson MW, Rudolph-Owen LA;
XX
DR WPI; 2003-646176/61.
DR N-PSDB; ADE38394.
XX
PT Treating subject having tumorigenic disorder or angiogenic disorder
PT caused by aberrant polypeptide e.g., N-formylpeptide receptor or nucleic
PT acid, by administering a modulator.
XX
PS Disclosure; SEQ ID NO 56; 454pp; English.
XX
CC This invention relates to a novel method of treating a human subject
CC having a tumorigenic disorder or angiogenic disorder, caused by aberrant
CC gene expression or activity of an isolated protein, by administering a

CC modulator. The modulator may have cytostatic, antithyroid, antidiabetic
CC or ophthalmological activity. The method is useful for treating a subject
CC having a tumorigenic or angiogenic disorder, in particular for treating
CC cancer (for example breast cancer, colon cancer, lung cancer or prostatic
CC cancer) and, for example, Grave's disease and diabetic retinopathy. The
CC present sequence is the amino acid sequence of the novel isolated human
CC protein 9389 of the invention.

XX
SQ Sequence 520 AA;

Query Match 98.8%; Score 2480; DB 7; length 520;
Best Local Similarity 91.9%; Pred. No. 8.1e-234;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MPGSLPLNABACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60
Db 1 MPGSLPLNABACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQAKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNNISYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNNISYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY 121 EGIIDTTNACYGCTAAVFNAVNWIESSWD----- 149
Db 121 EGIIDTTNACYGCTAAVFNAVNWIESSWDGRYALVVAGDIAYATGNARPTGVGAVALL 180
QY 150 -----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKI 198
Db 181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQTKASLLVSNQNGNMYTSSVYGLASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASSELFSQTKASLLVSNQNGNMYTSSVYGLASVLAQY 360
QY 319 SPQOLAGKRIGVFSYSGSLAATVLSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db 361 SPQOLAGKRIGVFSYSGSLAATVLSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVFAENMKLREDTHHLVNYIPOGSIDSLFEGTWYLVRVDEKHRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLVNYIPOGSIDSLFEGTWYLVRVDEKHRTYARRPTPNDDTLDEG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAAPEAAVISNGEH 478
Db 481 VGLVHSNIATEHIPSPAKKVPRLPATAAPEAAVISNGEH 520

Search completed: June 24, 2004, 13:17:51
Job time : 101 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2004, 06:24:23 ; Search time 4681 Seconds
(without alignments)
12771.648 Million cell updates/sec

Title: US-10-622-516-1

Perfect score: 2002
Sequence: 1 cgcctccacgcgactctcgg.....aaaaaaaaaaaaaaaa 2002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1191.4	59.5	3222	11 AK045094	AK045094 Mus muscu
2	1189.6	59.4	3286	11 AK044835	AK044835 Mus muscu
3	1189.6	59.4	3466	11 AK031297	AK031297 Mus muscu
4	829.4	41.4	960	12 BQ049460	BQ049460 AGENCOURT

5	800.6	40.0	881	12	BG697489	BG697489 602660682
6	788.6	39.4	1053	13	BU162936	BU162936 AGENCOURT
7	765.8	38.3	817	12	BG740730	BG740730 602631591
8	737	36.8	738	12	BG742704	BG742704 602632919
9	706.2	35.3	894	12	BG683150	BG683150 602651636
10	666.8	33.3	703	9	AU127534	AU127534 AU127534
11	664.2	33.2	1083	13	BX441322	BX441322 BX441322
12	657.6	32.8	795	12	BG683791	BG683791 602651636
13	644.2	32.2	733	12	BG698557	BG698557 602658594
14	643.6	32.1	789	14	CA512392	CA512392 UI-R-FJ0-
15	626	31.3	1201	13	BX419944	BX419944 BX419944
16	620	31.0	620	12	BM790716	BM790716 K-EST0070
17	618.2	30.9	779	14	CD354310	CD354310 UI-M-GM0-
18	611.8	30.6	780	14	CA316122	CA316122 UI-M-FW0-
19	608.4	30.4	998	12	BM474567	BM474567 AGENCOURT
20	601	30.0	601	10	AM249691	AM249691 2819708.3
21	596.4	29.8	914	13	BU941319	BU941319 AGENCOURT
22	596	29.8	950	14	CF412994	CF412994 CH3H083_D
23	595.8	29.8	937	13	BX441044	BX441044 BX441044
24	592	29.6	592	9	AL138300	AL138300 DKFP547K
25	580.2	29.0	682	14	CB288167	CB288167 CMD75_F05
26	576	28.8	810	14	CD354238	CD354238 UI-M-GM0-
27	573	28.6	725	14	CD354751	CD354751 UI-M-GM0-
28	572.6	28.6	905	13	BU508388	BU508388 AGENCOURT
29	572.2	28.6	681	14	CF169002	CF169002 B0807G11-
30	568.4	28.4	674	14	CF170711	CF170711 B0832B04-
31	568.4	28.4	877	9	AU132233	AU132233 AU132233
32	568.2	28.4	814	9	AU130614	AU130614 AU130614
33	568	28.4	718	12	BM963690	BM963690 UI-M-EQ0-
34	567.4	28.3	760	12	BG771780	BG771780 602720453
35	565.2	28.2	674	14	CA873594	CA873594 K0926C09-
36	561.8	28.1	669	14	CA874751	CA874751 K0934F10-
37	560.6	28.0	912	13	BQ720370	BQ720370 AGENCOURT
38	558.6	27.9	1059	12	BT463653	BT463653 603207318
39	552	27.6	657	14	CA884619	CA884619 K0934B05-
40	552	27.6	657	14	CA884619	CA884619 B0112E09-
41	552	27.6	1201	13	BX445244	BX445244 BX445244
42	551.4	27.5	555	10	AW951421	AW951421 EST363491
43	550.4	27.5	656	14	CA889355	CA889355 B0152D07-
44	550.4	27.5	657	14	CA892777	CA892777 B0174F04-
45	550	27.5	654	14	CF9000429	CF9000429 A0315D11-

ALIGNMENTS

RESULT 1
AK045094
LOCUS 3222 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length enriched library, clone:BI30032C06 product:pre B-cell leukemia transcription factor 1, full insert sequence.

ACCESSION AK045094
VERSION AK045094.1 GI:26337068
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE
AUTHORS
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kikunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishi,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3222)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishi,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akaira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. Tomohiro Kono (Department of Animal
Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
kanagawa Prefecture, Japan) whose assistance we gratefully
acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
location/Qualifiers
1. 3222
FEATURES
source
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB: B130032C06"
/db_xref="MGI: 2410453"
/db_xref="taxon:10090"
/clone="B130032C06"
/issue_type="parthenogenote"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9.5 days embryo"
62. 1624
/note="unnamed protein product; pre B-cell leukemia
transcription factor 1 (MGD|MGI:97495, GB|NM_008783,
evidence: BLASTN, 99%, match=124)

putative"
/codon_start=1
/protein_id="BAC32218.1"
/db_xref="GI:26337069"
/translation="MPGSLPLNAEACMPKDVIGVALEIYFSPQYVDOALEKYDGVDA
GKVTIGLQARMGFCTREDINSLCTLVVQKLMERSLSYDCIGRLVGTETIIDSK
SVKSNLQDLFEESGNDIEGIDTTNAGYGGTAVFNAVNWVESSNDGRYALVAGDI
AIYATGNARPTGCVGAVALLIGPNAPLIFDRGLRGTMHQAYDFYKPDMLSEYPVVG
KLISQCYLSALDRCYSVYRKIRAPQWKEGKDKDFTLNDGFMIFHSPYCKLVQKSLA
LTVSNQNGNMVTSVYGLASVLAQYSPQOLAGKRVGVSYGSGILATLYSLKVTQDA
TPGSALDKITIASLCDLKSRLDSRTGVAPDVFAENMKLRDTHLANIYIPQCSIDSLFE
GTWYLYVRVDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATEHIPBAKKVPRLPATSA
ESESAVISNGEH"
3208. 3213
/note="putative"
3222
/note="putative"
ORIGIN
polyA_signal
polyA_site
Query Match 59.5%; Score 1191.4; DB 11; Length 3222;
Best Local Similarity 79.0%; Pred. No. 9.7e-175;
Matches 1570; Conservative 0; Mismatches 271; Indels 146; Gaps 7;
89 TTTCGTGGCTCACTCCCTTTCCTCTGCTGCCGCTCGGTCAAGCTTGTCTTTCAACCATGC 148
Db TCTCCTTGCTTGTGCTGCTCTTCTTCTCCAGGCTGTGATCCCTTTGGCTCTTTCACCATGC 65
QY 149 CTGATCACTTCTTGAATGCAAGCTTGTGTCGCCCAAAAGATGTGGAAATGTTGCC 208
Db CTGGGTCACTTCTTGAATGCAAGGCTTGTGTCGCCCAAAAGATGTGGAAATGTTGCC 125
QY 209 TTGAGATTAATTTCTTCTCAATATGTGATCAAGCAGAGTTGAAAAATATGATGGTG 268
Db TTGAATCTAATTTCTTCTCAATATGTGATCAAGCTGAGTTGAAAAATACGATGGTG 185
QY 269 TAGATGCTGAAAGTATACCATTTGGCTTGGCCAGGCCAAGATGGCTTTCACAGATA 328
Db TAGATGCTGAAAGTATACCATCGGCTTGGCCAGGCCAAGATGGCTTTCACAGATC 245
QY 329 GAGAAATATTAACTCTCTTGCATGACTGTGTTCAGAATCTTATGAGAGAAATTAAC 388
Db GTGAAGACATCAACTCTCTTGCCTGACTGTGTTCAGAACCTGATGAGAGACATAGCC 305
QY 389 TTTCTATGATTGCTTGGCGGCTGGAAGTTGGAACAGAGACAATCATCGACAATCAA 448
Db TTTCTATGATTGCTTGGCGGCTGGAAGTTGGAACAGAGACAATCATCGACAATCGA 365
QY 449 AGTCTGTAAGACTAATTGATGACGCTGTTGAAAGAGTCTGGAATACAGATATTAAG 508
Db AATCAGTGAAGTCTAATTGATGACGCTGTTGAGAGAGTCTGGAATACAGATATTAAG 425
QY 509 GAATCGACACAATAATGATGCTATGAGAGACAGCTGCTGCTTCAATGCTGTTAACT 568
Db GAATGATACACAATGATGCTATGAGAGAGAGTCTGGAATACAGATATTAAGAACT 485
QY 569 GAATTGAGTCCAGCTCTTGGGAT----- 591
Db GGGTGAATCCAGCTCTTGGGATGAGCAGATATGCTGTGTAAGTTGACAGAGACATTGCTA 545
QY 592 ----- 591
Db TATATGCCACAGAAATGCCAGACCTACAGGTGAGATTGGAAGCTGTGGCCTGCTAATTG 605
QY 592 -----GGGCTTGTGGGACACACATATGCAACATGCTT 622
Db GGCACAAAGCTCTTAATTTTTCACGAGGGCTCCGTGGGACACACATGACATGCTT 665
QY 623 ATGATTTTACAAGCGCTGATATGCTATGATATATCTTATAGTAGAGAAAACCTCTTCA 682
Db ATGACTTTTACAAGCGCTGATATGCTATGATATATCTTATAGTAGAGAAAACCTCTTCA 725
QY 683 TACAGTCTAAGCTGATGATGACCGCTGCTACTGTCTACTGCAAAAAGATCATTG 742

Db	726	TACAGTGCCTACGCGCCCTGGACCGCTGCTATCTGTCTACCGCAAAAAGATCCGTG	785
Qy	743	CCCACTGGCAGAAAGAGGAATGATAAGATTTTACCTGAATGATTTTGGCTTCATGA	802
Db	786	CCCACTGGCAGAAAGAGGGAAGGATTAAGATTTTACCCTGAATGATTTTGGCTTCATGA	845
Qy	803	TCTTCACTCACCATATTGTAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTGAATG	862
Db	846	TCTTCACTCACCATATTGTAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTGAATG	905
Qy	863	ACTTCCCTTAATGACAGAAATAGAGATAAAAAATAGTATCTATAGTGGCCCTGGAAGCTTTG	922
Db	906	ACTTCTTAATGATCAAAAACAGAGATAAAAAACAGTATTTTACAGTGAGCTGGAAGCTTTG	965
Qy	923	GGGATGTTAAATTTAGAGACACCTACTTTGATAGAGATGTGGAAGGCAATTATGAAGG	982
Db	966	GGGACGTTAAATTTAGAGATACTTACTTTGACAGAGATGTAGAAAAAGCATTTATGAAGG	1025
Qy	983	CTAGCTCTGAACCTTCAGTCAAGAAAACAAAGGCATCTTTACTTGATCAAAATCAAAATG	1042
Db	1026	CTAGTCTGAGCTATTTCAACAGAAAACAAAGGCGTCTTGCTGTGTCTAATCAGATG	1085
Qy	1043	GAAATATGTACACATCTTCAGTATATGTTCCCTTGCATCTGTTCTAGCACAGTACTCAC	1102
Db	1086	GAAATATGTACACATCTCTGTCTATGTTCCCTTGGCTTCTGTCTTGCCAGACAGTACTCAC	1145
Qy	1103	CTCAGCAATTAGCAGGGAAGAGAAATTGAGTGTCTTATAGTCTGTGTTGGCTGCCA	1162
Db	1146	CTCAGCAATTGCGAGGGAAGAGGTTGAGTGTCTTATAGTCTGTGTTGGCTGCCA	1205
Qy	1163	CTCTGTACTCTCTTAAAGTCAACAAGATGCTACACGCGGCTGCTCTTGATTAATAA	1222
Db	1206	CGCTGTACTCCCTTAAAGTCAACAAGACGCCACACAGGATCTGCCCTTGATTAATAA	1265
Qy	1223	CAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGTTGGCAGCAGATG	1282
Db	1266	CAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGTTGGCAGCAGATG	1325
Qy	1283	TCTTCGCTGAAAACATGAAGCTCAGAGAGGACCCCATCATTTGGTCAACTATATTC	1342
Db	1326	TCTTCGCTGAAAACATGAAGCTCAGAGAGGACCCCATCATTTGGTCAACTATATTC	1385
Qy	1343	AGGGTCAATAGATTCACTTTTGAAGGAACGTGGTACTTAGTTAGGTTGATGAAAAAC	1402
Db	1386	AGTGTCAATAGATTCACTTTTGAAGGAACGTGGTACTTAGTTAGGTTGATGAAAAAC	1445
Qy	1403	ACAGAAAGACTTAACGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAGGATAG	1462
Db	1446	ACAGAAAGACTTAACGCGCGCGCCCTTCACAAATGACACAGTTGGATGAAGGATAGG	1505
Qy	1463	GACTTGTGATTCAAAACATAGCAACTGAGCATATTTCCAAAGCCCTGCCAAGAAAGTACCA	1522
Db	1506	GGCTCGTGCATAGTAACACAGCAACAGAGCATATTTCCAAAGCCCTGTAAGAAAGTACCA	1565
Qy	1523	GACTCCCTGCACAGCAGAGAACTGAAAGCAGCTGTCAATTAGTATGGGGAACATTAAAG	1582
Db	1566	GACTCCCTGCACACCTCGGCCGAATCTGAATCAGCTGTCACTAAGTAAAGGGAAGCACTGAG	1625
Qy	1583	ATACTCTGTGAGGTGCAAGACTTCAAGGTTGGGTTGGGTCATGGGGTATGGGAACA	1642
Db	1626	AGTCTGTGGCTTTACA-----GAGGCTCGGGGCGGGGATGGGGTATGGGAACG	1674
Qy	1643	GTTGAGGAATGGGATATCTGGGGAATAATTAAAGATTAACATGTTATGTAAATTTTAA	1702
Db	1675	GTTGAGGAAT--GGATGTCTTGGGAACAATTTTGA--GATTAATGTGTCTTAAATGTAA	1732
Qy	1703	TGTGACTGACATGAGCCTGGAATGACTATCGTGTACTTGGGAAGTCT--CTTGTCTCT	1759
Db	1733	TGTAATGACACAGAGCCGAGAAAGCTATTTGTTTGGAAAAAGTCTCTGCTGAATGAA	1792
Qy	1760	ATTGCTGACATGCTTCTGTGTGTGTCTGGCCAATGCAAAATGTACTGAAATGATGTT--	1818

Db	1793	ACTTGCTAACATGCTTCCTCTGTGTGTGCTCTGGCCAGTGTAAATGTACTGCAGTGATGTTA	1852
Qy	1819	AAGGGCTCTGTAAAACTTCATACCTCTTTGGCCATTGTATGATGATGTTGGTTTTA	1878
Db	1853	AAGGCTCTGTAGAACTTTATACCTCTCTGGATGTTTATACGCATGACGATTAGTTTCA	1912
Qy	1879	AACATGGTATATAATGAATTGTGTACTTCTGTGCAGAGAAAGCAGAGTACTTAATCTCCAAT	1938
Db	1913	GATGTGTT--ATGAACCTGAGTGTCTTGTGACGCAAAAGCAGAGGTAAGTGTGTCAT	1969
Qy	1939	TAAAAAA	1945
Db	1970	TTTTTAA	1976
RESULT 2			
AK044835			
LOCUS			
DEFINITION	AK044835	3286 bp	mRNA
			linear
			HTC 20-SEP-2003
			Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
			enriched library, clone:Bl3007E24 product:pre B-cell leukemia
			transcription factor 1, full insert sequence.
ACCESSION	AK044835		
VERSION	AK044835.1	GI:263336856	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1		
AUTHORS	Carninci,P. and Hayashizaki,Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,		
	Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
	Normalization and subtraction of cap-trapper-selected cDNAs to		
	prepare full-length cDNA libraries for rapid discovery of new genes		
	Genome Res. 10 (10), 1617-1630 (2000)		
JOURNAL	20499374		
MEDLINE	11042159		
PUBMED			
REFERENCE	3		
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,		
	Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,		
	Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,		
	Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,		
	Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,		
	Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,		
	Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.		
	RIKEN integrated sequence analysis (RISA) system--384-format		
	sequencing pipeline with 384 multipicapillary sequencer		
	Genome Res. 10 (11), 1757-1771 (2000)		
JOURNAL	20530913		
MEDLINE	11076861		
PUBMED			
REFERENCE	4		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the		
	FANTOM Consortium.		
	Functional annotation of a full-length mouse cDNA collection		
	Nature 409, 685-690 (2001)		
TITLE	5		
JOURNAL	The FANTOM Consortium and the RIKEN Genome Exploration Research		
REFERENCE	Group Phase I & II Team.		
AUTHORS	Analysis of the mouse transcriptome based on functional annotation		
	of 60,770 full-length cDNAs		
	Nature 420, 563-573 (2002)		
TITLE	6 (bases 1 to 3286)		
JOURNAL	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,		
REFERENCE	Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,		
AUTHORS	Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,		
	Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,		
	Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,		
	Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,		

Qy	1208	CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTTCAAGAACTG	1267
Db	1312	CCCTTGATAAAATAACAGCAAGTTTATGTGACCTTAAATCAAGGCTTGACTCGAAGACTT	1371
Qy	1268	GTGTGGCACCAAGATGTCCTCGCTGAAAACATGAAGCTCAGAGAGACACCATCATTTGG	1327
Db	1372	GTGTGGCACCCGATGTCCTTGCTGAAAACATGAAGCTCAGAGAGACACATCATTTAG	1431
Qy	1328	TCAACTATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGAACTGTACTTATGTTA	1387
Db	1432	CCAACATATATTTCCCAAGTGTCAATAGACTCACTCTTTGAAGAACTGTATCTGTCA	1491
Qy	1388	GGGTGATGAAAAGCACAGAGAAGACTTACGCTCGGCGTCCCACTCCAATGATGACACTT	1447
Db	1492	GAGTGATGAAAACACAGAGAAGACTTACGCCCGCGCTTCAAAATGACCAAGACTT	1551
Qy	1448	TGGATGAAGAGTAGGACTTGTGCATTCAAAACATAGCAACTGAGCATATTTCCAAGCCTG	1507
Db	1552	TGGATGAAGGAATGGGGCTCGTGCAATAGTAACACAGAACAGACATATTTCCAAGCCTG	1611
Qy	1508	CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCAATAGTA	1567
Db	1612	CTAAGAAAGTGCCAAGACTCCCTGCCAACCTCGGCCGAATCTGAATCAGCTGTCAATAGTA	1671
Qy	1568	ATGGGAAACATTAAAGTACTCTGTGAGGTGCAGAAGCTTCAAGGTGGGTGGGCATGGCGT	1627
Db	1672	ACGGGAGCACTGAGAGTCTGTGGCCTTTACA-----GAGGCTCGGGGCGGGGA	1720
Qy	1628	GGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGATATTTTAAAGGATTACATG	1687
Db	1721	TGGGGTATGAAACGGTTGAGGAAT-GGATGCTTTGGGACAAATTTTGCA-GATTATGTG	1778
Qy	1688	TTATGTAAATTTTATGTGACTGACATGAGCGCTGGATGACTATCGTGTACTTTGGAAAG	1747
Db	1779	TTGCTTAAATGTATGTAACTGACACAGAGCCAGAAAGCTATTTGTGTTTGGAAAG	1838
Qy	1748	TCT--CTTGTCTATTTGCTGACATGCTTCTGTTGTGTCTGGCCAATGCCAAATGT	1804
Db	1839	TCTCTGCTGAACCTGAACCTGTAAACATGCTTCTGTTGTGTCTGGCCAGTGTAAATGT	1898
Qy	1805	ACTGGAATGATGTT-AGGGCTCTGTAAAACTTCATACCTTTTGCCATTGTATGCAT	1863
Db	1899	ACTGCAAGTATGTTAAAGGGCTCTGTAGAACTTATACCTCTCTGGATGTTATACGAT	1958
Qy	1864	GATGTTTGGTTTAAACATGGTATATGAATTGTACTTCTGTCAAGAGAAGACAGAG	1923
Db	1959	GCAGTTTAGTTTTCAGATGTGCT--ATGAAGTGAAGTCTTGTACAGCAAAAGGCAGAG	2015
Qy	1924	GTAATAATCTCCAATTAATAA	1945
Db	2016	GTAATAAGTGTCCAATTTTAA	2037
RESULT 3			
LOCUS	AK031297	3466 bp	mRNA linear HTC 18-SEP-2003
DEFINITION	Mus musculus 13 days embryo male testis cDNA, RIKEN full-length enriched library, clone:6030403N11 product:pre B-cell leukemia transcription factor 1, full insert sequence.		
ACCESSION	AK031297		
VERSION	AK031297.1	GI:26327188	
SOURCE	HTC; CAP trapper.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		

AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL MEDLINE PUBMED	Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159
REFERENCE	3
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL MEDLINE PUBMED	Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL MEDLINE PUBMED	Nature 409, 685-690 (2001) 20530913 11076861
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL MEDLINE PUBMED	Nature 420, 563-573 (2002) 20530913 11076861
REFERENCE	6 (bases 1 to 3466)
AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohnato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL MEDLINE PUBMED	Submitted (16-JUL-2001)
REFERENCE	Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/.
FEATURES	location/Qualifiers 1. .3466 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:6030403N11" /db_xref="MGI:2354508" /db_xref="taxon:10090" /clone="6030403N11" /sex="male" /tissue_type="testis" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="13 days embryo"

CDS
122..1684
/note="unnamed protein product; pre B-cell leukemia
transcription factor 1 (MGD|MG1:97495, GB|NM_008783,
evidence: BLASTN, 99%, match=124)
putative"
/codon_start=1
/protein_id="BAC27338.1"
/db_xref="GI:26327189"
/translation="MPGSLPINAECWPKDVGIVALEIYFSPQYVDQAELEKYDGVDA
GKXTIGLGQARMGFCTDREDINSLCTIVQKMERHSLSYDCIGRLVGTETIIDSK
SVKSNLMQLFEESGNTDIEGIDFTNACYGGTAAYENAVNWNWESSMDGRYALVAVGDI
AIVATGNARPTGVGVAVALLIGNAPLIFDRGLRGTMOHAYDFYKPDMLSEYPVVDG
KLSTQCYLSALDRCYSVYRKIRIAQWQKEGKDFTLNDFGHMFHSPYCKLVQKSLA
RMFLNDFLNDONRDKNSIYSGLEAFDVKLEDTYFDRDVEKAFMKASSELFNQKTAS
LIVSNQNGNMVTSVYSGSLASVLAQYSPQOLAGKRVGVFYSYSGLAATLYSLKVTODA
TPGSALDKITASLCDLKSRLDSRTCVAPDVFENMKLRDETHLANIIPQCSIDSLFE
GTMVYLVRVDEKHKRRTYARRPFTNDHSLDEGMGLVHSNTATEHIPSAPKVRRLPATSA
ESESAVISNGEH"
3447..3452
/note="putative"
3466
/note="putative"
polya_signal
polya_site
ORIGIN
/note="putative"

Query Match 59.4%; Score 1189.6; DB 11; Length 3466;
Best Local Similarity 79.9%; Pred. No. 1.8e-174;
Matches 1552; Conservative 0; Mismatches 244; Indels 146; Gaps 7;

QY 134 GCTCTTCAACCATGCTCGATCACTTCTTTGAATGCAGAAAGCTTGCGCCAAAGATG 193
Db 111 GCTCTTCAACCATGCTCGGATCACTTCTTTGAATGCAGAGGCTTGCGCCAAAGATG 170
QY 194 TGGGAATGTGCTGAGATCTATTTCTTCTCAATATGTGATCAAGCAGAGTTGG 253
Db 171 TGGGAATGTGCTGAGATCTATTTCTTCTCAATATGTGATCAAGCAGAGTTGG 230
QY 254 AAAAATATGATGTGTAGATGTGGAAGATATACCATTTGGCTGGCCAGGCCAAGATGG 313
Db 231 AAAAATATGATGTGTAGATGTGGAAGATATACCATTTGGCTGGCCAGGCCAAGATGG 290
QY 314 GCTCTGACAGATAGAGAAGATATTACTCTTTTGCATGACTGTGTGCAATCTTA 373
Db 291 GCTCTGACAGATCGTGAAGATCAACTCTTTTGCATGACTGTGTGCAATCTTA 350
QY 374 TGGAGAGAAATACTTTCTTATGATTCATTTGGCGCGCTGGAAGTTGGAACAGACAA 433
Db 351 TGGAGAGACATAGCTTTCTTATGATTTGATGGCGCGCTGGAAGTTGGAACAGACAA 410
QY 434 TCATCGACAATCAAAAGTCTGTGAAGCTAAATTGATGACGCTGTTGAAGTCTGGGA 493
Db 411 TCATCGACAATCGAAATCAAGTGAAGTCTAAATTGATGACGCTGTTGAAGTCTGGGA 470
QY 494 ATACAGATATAGAGAAGATGCACAACCTAATGCATGCTATGAGGCAAGCTGCTGTCT 553
Db 471 ATACAGATATAGAGAAGATGCACAACCTAATGCATGCTATGAGGCAAGCTGCTGTCT 530
QY 554 TCAATGCTGTAACTGATGATCCAGCTCTTGGGAT----- 591
Db 531 TCAATGCTGTAACTGATGATCCAGCTCTTGGGATGACGATATGCTGTGTAGTTG 590
QY 592 ----- 591
Db 591 CAGAGACATTGCTATATATGCCACAGAAATGCCAGACCTTACAGTGGAGTTGAGCTG 650
QY 592 -----GGGCTTGGTGGGACAC 607
Db 651 TGGCCCTGCTAATTGGGCCAAACGCTCTTAATTTTGAACGAGGCTCCGTGGACAC 710
QY 608 ATATGCAACATGCTATGATTTTACAAGCTGATATGCTATCTGAATATCTTATAGTAG 667
Db 711 ACATGCAACATGCTATGATTTTACAAGCTGACATGCTCTCCAGATAACCTGTGTGCG 770
QY 668 ATGAAAACTCTCATATACAGTGTACTCTAGTGCATTAGAACCGCTGCTACTCTGTCTACT 727

Db 771 ACGGAAAGCTCTCCATACAGTGTCTACCTCAGCGCCCTGGACCGCTGCTATTTCTGTCTACC 830
QY 728 GCAAAAAGATCCATGCCAGTGGCAGAAAGAGGAAATGATAAGATTTTACCTTGAATG 787
Db 831 GCAAAAAGATCCGTGCCAGTGGCAGAAAGAGGAAATGATAAGATTTTACCTTGAATG 890
QY 788 ATTTGGCTTCATGATCTTTCACTCAACCATATTTGTAACCTGGTTCAAGAAATCTTACGCTC 847
Db 891 ATTTGGCTTCATGATCTTTCACTCAACCATATTTGTAACCTGGTTCAAGAAATCTTACGCTC 950
QY 848 GGATGTGCTGAATGACTTCTTAATGACCAAGATAGATAAATAGTATCTATAGTG 907
Db 951 GGATGTGCTGAATGACTTCTTAATGATCAAAAAGAGATAAAAAAGTATTTACAGTG 1010
QY 908 GCGTGAAGCCTTTGGGAGTGTAAATTAGAGACACCTACTTTGATAGAGATGTGAGA 967
Db 1011 GACTGGAAGCCTTTGGGAGCTTAAATTAGAGATACTTACTTTGACAGAGATGTAGAAA 1070
QY 968 AGGCATTATGAAGGCTAGCTGTGAACCTCTCAGTCAGAAAACAAAGGCATCTTACTTG 1027
Db 1071 AGGCATTATGAAGGCTAGCTGTGAACCTCTCAGTCATCAACCAAGAAACAAAGGCCTTGTCTTG 1130
QY 1028 TATCAATCAAAATGGAATATGTACACATCTTCAATATAGTTCCTTGCATCTGTTG 1087
Db 1131 TGTCTAATCAGATGGAATATGTACACATCTCTGTATAGTTCCTTGTCTGTCC 1190
QY 1088 TAGCACAGTACTCACTCAGCAATTAAGCAGGAGAAAGATTGAGTGTCTTATGTT 1147
Db 1191 TGGCACAGTACTCACTCAGCAATTAAGCAGGAGAAAGGTTGAGTGTCTTATCGTT 1250
QY 1148 CTGTTGGCTGCCACTCTGTACTCTTTAAAGTCAACAAGATGCTACACCGGGTCTG 1207
Db 1251 CTGTTGGCTGCCACTCTGTACTCTTTAAAGTCAACAAGATGCTACACCGGGTCTG 1310
QY 1208 CTCTGTAAATAATACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAGAAAGCTG 1267
Db 1311 CCCTGTAAATAATACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAGAAAGCTT 1370
QY 1268 GTGTGACACAGATGTCTGCTGGAACATGAAGCTCAGAGAGACACCATCATTTGG 1327
Db 1371 GTGTGACACCGATGTCTTGTGTAACATGAAGCTCAGAGAGACACATCATTTAG 1430
QY 1328 TCACTATATTTCCCAAGGTTCAATAGATTCACTTTTGAAGAAAGTGTACTTATG 1387
Db 1431 CCACTATATTTCCCAAGGTTCAATAGATTCACTTTTGAAGAAAGTGTACTTATG 1490
QY 1388 GGTGTGATGAAAAGACAGAAAGCTTACGCTGGCGTCCCACTCCAAATGATGACACTT 1447
Db 1491 GAGTGTGATGAAAAGACAGAAAGCTTACGCTGGCGTCCCACTCCAAATGATGACACTT 1550
QY 1448 TGGATGAAGAGTATGAGACTTGTGCAATTCGAACATGACATATTCAGAGCCCTG 1507
Db 1551 TGGATGAAGAGTATGAGAGCTCGTGCATATGATACACAGCAACAGACATATTCAGAGCCCTG 1610
QY 1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGACAGAACTGAAAGCAGCTGTCTATTA 1567
Db 1611 CTAAGAAAGTACCAAGACTCCCTGCCACAGACAGAACTGAAAGCAGCTGTCTATTA 1670
QY 1568 ATGGGGAACATTAAATGATCTGTGAGTGAAGAACTTCAAGGTTGGGTGGGATGGGT 1627
Db 1671 ACGGGGAGCACTGAAGTCTGTGGCTTTACA-----GAGGCTCGGGGCCGGGA 1719
QY 1628 GGGGCTATGGGAAACAGTTGGAAGATGGGATATCTGGGGAATTTTAAAGATTTACATG 1687
Db 1720 TGGGCTATGGAACAGTTGGAAGAAAT-GGATGTCTTGGGACAAATTTTGCA-GATTATGTG 1777
QY 1688 TTATGTAAATTTTATGTAAGTGAATGAGAGCTGATGATCTATCTGTACTTGGGAAAG 1747
Db 1778 TTGCTTAAATGTATGTATGTAAGTGAACAGAGCCCAAGAACTATTTGTGTTTGGAAAG 1837
QY 1748 TCT--CTTGTCTAATTTGCTGACATGCTCTGTTGTTGTTGCTGGCCCAATGCT 1804

Db 1838 TCTCTGCTGAACCTGAACCTTGCTAACAATGCTTCTCTGTGTGCTG6CCAGTGTAATGT 1897

Qy 1805 ACTCGAATGATGT-AAGGCTCTGTAAACTTCATACCTCTTTGGCCATTGTATGAT 1863

Db 1898 ACTGCAGTATGTTAAAGGGCTCTGTAGAACCTTATACCTCTCGATGTTATACGCAT 1957

Qy 1864 GATGTTTGGTTTTTAAACATGATATGAATTGTACTTCTGTCAAGAAAGCAGAG 1923

Db 1958 GCAGTTTAGTTTTCAGATGTGGT--ATGAACGTAGTGTCTGACAGCAAAAGGCAGAG 2014

Qy 1924 GTACTAATCTCCAATTAAAAA 1945

Db 2015 GTACTAGTGTCCAATTTTAA 2036

RESULT 4

BQ049460 960 bp mRNA linear EST 29-MAR-2002

LOCUS AGENCOURT_7064727 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5787318

DEFINITION 5', mRNA sequence.

ACCESSION BQ049460

VERSION BQ049460.1 GI:19808800

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue Procurement: Lou Staudt cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov plate: LLM12877 row: h column: 07 High quality sequence stop: 595.

FEATURES

Source 1..960

location/Qualifiers

1..960

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5787318"

/cissue_type="lymphoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 85"

/note="Organ: lymph; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 41.4%; Score 829.4; DB 12; Length 960;

Best Local Similarity 94.2%; Pred. No. 1.2e-118;

Matches 905; Conservative 0; Mismatches 51; Indels 5; Gaps 4;

Qy 865 TTCCTTAATGACCAAGATAGATTAATAATAGTATCTATAGTGGCCTGGAAGCCTTTGGG 924

Db 1 TTCCTTAATGACCAAGATAGATTAATAATAGTATCTATAGTGGCCTGGAAGCCTTTGGG 60

Qy 925 GATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGAGGCATTATGAAGGCT 984

Db 61 GATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGAGGCATTATGAAGGCT 120

Qy 985 AGCTCTGAACCTCTTCAGTCAGAAACAAGGCATCTTACTTGTATCAAAATCAAAATGGA 1044

Db 121 AGCTCTGAACCTCTTCAGTCAGAAACAAGGCATCTTACTTGTATCAAAATCAAAATGGA 180

Qy 1045 AATATGTACACATCTTCAGTATATAGTTCCTCCCTTGCAATCTGTTCTAGCACAGTACTCACT 1104

Db 181 AATATGTACACATCTTCAGTATATAGTTCCTCCCTTGCAATCTGTTCTAGCACAGTACTCACT 240

Qy 1105 CAGCAATTAGCAGGGAAGAAATTGGAGTGTTCCTTATGCTTCGTTGGCTGCCACT 1164

Db 241 CAGCAATTAGCAGGGAAGAAATTGGAGTGTTCCTTATGCTTCGTTGGCTGCCACT 300

Qy 1165 CTGTACTCTCTTAAAGTCACAAGAATGCTACACCCGGGCTGCTCTTGATTAATAATACA 1224

Db 301 CTGTACTCTCTTAAAGTCACAAGAATGCTACACCCGGGCTGCTCTTGATTAATAATACA 360

Qy 1225 GCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAAGAACTGCTGGCACCAGATGTC 1284

Db 361 GCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAAGAACTGCTGGCACCAGATGTC 420

Qy 1285 TTCGCTGAAAACATGAAGCTCAGAGAGACACCCATCAATTGGTCAACTATATTCGCCAG 1344

Db 421 TTCGCTGAAAACATGAAGCTCAGAGAGACACCCATCAATTGGTCAACTATATTCGCCAG 480

Qy 1345 GGTTCATATAGATTCACTCTTTGAAGGAACGTGTTAGTTAGGGTGATGAAAAGCAC 1404

Db 481 GGTTCATATAGATTCACTCTTTGAAGGAACGTGTTAGTTAGGGTGATGAAAAGCAC 540

Qy 1405 AGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACA-CTTTGGATGAAGAGATAGG 1463

Db 541 AGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAGAGATAGG 600

Qy 1464 ACTGTGCAATCAAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTACCAAG 1523

Db 601 ACTGTGCAATCAAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTACCAAG 660

Qy 1524 ACTCCCTGCCACAGCAGCAAGAACTGAAGCAGCTGTCAATTAGTAATGGGAACATTAGA 1583

Db 661 ACTCCCTGCCACAGCAGCAAGAACTGAAGCAGTGTCAATTAGTAATGGGAACATTAGA 720

Qy 1584 TACT-CTGTAGGTGCAAGACTTCAGGGTGGGTGGG--ATGGGGTGGGGTATGGGA 1640

Db 721 TACTCTGGAGGTGCAAGACTTCCGGGTGGGGGGGCCATGGGGGGGGTATGGGGA 780

Qy 1641 CAGTTGAGGAATGGGATATCTGGGGATTAATTTAAAGATTACATGTTATGTAATTTT 1700

Db 781 CAGTTGAGGAATGGGATATCT-GGGAAAAATTTTAAAGATTACCGGGTAGGGAATTTT 839

Qy 1701 TATGTACTGACATGGAGCCTCGATGACTATCGTACTTGGAAAAAGTCTTTGCTCTA 1760

Db 840 TATGTACTGACAGAGGAGCCGGATGACTAACCGGAGCTGGGAAAAAGCCCTTGCCCA 899

Qy 1761 TTTCGTGACATGCTTCTGTGTGTCTGGCCAATGCCAATGTACTCGAATGATGTTAA 1820

Db 900 TTTCGCGGAGCGCTTCGGGGGGGGGCTGGCCAAGGCCAAGGTATCCAAGAGAGCAA 959

Qy 1821 G 1821

Db 960 G 960

RESULT 5

BG697489 881 bp mRNA linear EST 07-MAY-2001

LOCUS BG697489

DEFINITION 60260682F1 NCI_CGAP_skn3 Homo sapiens cDNA clone IMAGE:4803788 5', mRNA sequence.

ACCESSION BG697489

VERSION BG697489.1 GI:13963764

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10700 row: c column: 21
High quality sequence stop: 815.

FEATURES
source Location/Qualifiers
1..881
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4803788"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Skin3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP library."

ORIGIN

Query Match 40.0%; Score 800.6; DB 12; Length 881;
Best Local Similarity 97.4%; Pred. No. 3.5e-114;
Matches 856; Conservative 0; Mismatches 19; Indels 4; Gaps 4;
OY 964 GAGAGGCATTATGAGGCTAGCTCTGAACTCTTCAGTCAGAAACAAAGGCATCTTTA 1023
Db 1 GAGAGGCATTATGAGGCTAGCTCTGAACTCTTCAGTCAGAAACAAAGGCATCTTTA 60
OY 1024 CTGTATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGGTCCCTGCATCT 1083
Db 61 CTGTATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGGTCCCTGCATCT 120
OY 1084 GTTCTAGACAGTACTCACCCTCAGCAATTAGCAGGGAAGAAATGGAGTCTTTCTTAT 1143
Db 121 GTTCTAGACAGTACTCACCCTCAGCAATTAGCAGGGAAGAAATGGAGTCTTTCTTAT 180
OY 1144 GGTCTGGTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGG 1203
Db 181 GGTCTGGTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGG 240
OY 1204 TCTGCTTGATAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGA 1263
Db 241 TCTGCTTGATAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGA 300
OY 1264 ACTGATGGCACCAAGATGCTCTCGCTGAAACATGAAGCTCAGAGAGACCCCATCAT 1323
Db 301 ACTGATGGCACCAAGATGCTCTCGCTGAAACATGAAGCTCAGAGAGACCCCATCAT 360
OY 1324 TTGGTCAACTATATCCCCAGGGTCAATAGATTCACCTTTGAAGGAACGTGTAATT 1383
Db 361 TTGGTCAACTATATCCCCAGGGTCAATAGATTCACCTTTGAAGGAACGTGTAATT 420
OY 1384 GTTAGGGTGATGAAAGCAGAGAAGACTTAAGCTCGGCGTCCCACTCCAAATGATGAC 1443
Db 421 GTTAGGGTGATGAAAGCAGAGAAGACTTAAGCTCGGCGTCCCACTCCAAATGATGAC 480
OY 1444 ACTTTGATGAGAGAGTAGACTTGTGCATTC-AAACATAGCAACTGAGCATATTCAG 1502
Db 481 ACTTTGATGAGAGAGTAGACTTGTGCATTCAAACATAGCAACTGAGCATATTCAG 540
OY 1503 CCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACTGAAGCAGCTGTCA 1562
Db 541 CCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACTGAAGCAGCTGTCA 600
OY 1563 TAGTAATGGGAAACATTAAAGTACTCTGTAGGTGCAAGACTTCAGGGTGGGTGGGCAT 1622
Db 601 TAGTAATGGGAAACATTAAAGTACTCTGTAGGTGCAAGACTTCAGGGTGGGTGGGCAT 660

OY 1623 GGGTGGGGGTATGGGAACAGTTGGAGGAAT-GGGAATCTGGGGATATTTTAAAGAT 1681
Db 661 GGGTGGGGGTATGGGAACAGTTGGAGGAATGGGAATATCTGGGATATTTTAAAGAT 720
OY 1682 TA-CATGTTATGTAATTTTATGTGACTGACATGGAGCCTGGATGACTATCGTGACTT 1740
Db 721 TACCATGTGATGTCATTTTATGTGACTGACATGGAGCCTGGATGACTATCGTGACTT 780
OY 1741 GGAAGTCTCTTGGCTCTATTGGCTGACATGCTTCTGTGTGTGGCCAAATGCCAA 1800
Db 781 GGAAGTCTCTTGGCTCTATTGGCTGACATGCTTCTGTGTGGAGACTGGCC-ATGCCAA 839
OY 1801 ATGTACTCGAATGATGTTAAGGGCTCTGTAAACTTCAT 1839
Db 840 ATGTACTCGAAGGCTGTTAAGGGCTGTGAAAAATTTAT 878

RESULT 6
BU162936 1053 bp mRNA linear EST 04-SEP-2002
LOCUS AGENCOURT_6911057 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5952255
DEFINITION 5', mRNA sequence.
ACCESSION BU162936
VERSION BU162936.1 GI:22676846
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 1053)
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM2139 row: p column: 16
High quality sequence start: 3
High quality sequence stop: 693.

FEATURES
source Location/Qualifiers
1..1053
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5952255"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_110"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

ORIGIN

Query Match 39.4%; Score 788.6; DB 13; Length 1053;
Best Local Similarity 94.5%; Pred. No. 2.3e-112;
Matches 861; Conservative 0; Mismatches 44; Indels 6; Gaps 4;
OY 896 GTATCTATAGTGCCCTGGAAGCCTTTGGGATGTTAAATTAGAAGACACCTACTTGATA 955
Db 5 GTGGCAATTCCGACGAGGAAGCCTTTGGGATGTTAAATTAGAAGACACCTACTTGATA 64

QY 956 GAGATGTGAGAAAGCATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAAGG 1015
|||||
Db 65 GAGATGTGAGAAAGCATTTATGAAGGCTAGCTCTGAACCTTTCAGTCAGAAAACAAAGG 124

QY 1016 CATCTTTACTTGTATCAAAATGGAATATGTACACATCTTCAGTATATGGTCCC 1075
|||||
Db 125 CATCTTTACTTGTATCAAAATGGAATATGTACACATCTTCAGTATATGGTCCC 184

QY 1076 TTGCATCTGTTCTAGACACAGTACTCACCTCAGCAATTAGCAGGAAGAAATTGAGTGT 1135
|||||
Db 185 TTGCATCTGTTCTAGACACAGTACTCACCTCAGCAATTAGCAGGAAGAAATTGAGTGT 244

QY 1136 TTTCTTATGTTCTGTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACACAGATGCTA 1195
|||||
Db 245 TTTCTTATGTTCTGTTGGCTGCCACTCTGTACTCTCTTAAAGTCAACACAGATGCTA 304

QY 1196 CACCGGGGCTGCTCTTGATATAAATAACAGCAAGTTATGTGATCTTAAATCAAGGCTTG 1255
|||||
Db 305 CACCGGGGCTGCTCTTGATATAAATAACAGCAAGTTATGTGATCTTAAATCAAGGCTTG 364

QY 1256 ATTCAGAACTGTGTGGCAGCAGATGTCTTCGCTGAATAACATGAAGCTCAGAGAGACA 1315
|||||
Db 365 ATTCAGAACTGTGTGGCAGCAGATGTCTTCGCTGAATAACATGAAGCTCAGAGAGACA 424

QY 1316 CCCATCATTTGGTCAACTATATTTCCCAAGGGTTCATATAGATTCACTCTTTGAAGAACGT 1375
|||||
Db 425 CCCATCATTTGGTCAACTATATTTCCCAAGGGTTCATATAGATTCACTCTTTGAAGAACGT 484

QY 1376 GGTACTTAGTTAGGGTGGATGAAAAACACAGAAAGACTTACGCTGGCGTCCCACCTCAA 1435
|||||
Db 485 GGTACTTAGTTAGGGTGGATGAAAAACACAGAAAGACTTACGCTGGCGTCCCACCTCAA 544

QY 1436 ATGATGACACTTTGGATGAAGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATTA 1495
|||||
Db 545 ATGATGACACTTTGGATGAAGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATTA 604

QY 1496 TTCCAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGAGAAGCTGAAGCAG 1555
|||||
Db 605 TTCCAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGAGAAGCTGAAGCAG 664

QY 1556 CTGTCAATTAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGG 1615
|||||
Db 665 CTGTCAATTAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGG 724

QY 1616 TGGGCAT-GGGGTGGGGGTATGGGAACAAGTTGAGGAATGGGATATCTGGGATTAATTT 1674
|||||
Db 725 TGGGCATGGGGGTGGGGGTATGGGAACAAGTTGAGGAATGGGATATCTGGGATTAATTT 784

QY 1675 AAAGGATTACATGTTAT-GTAAATTTTATGTGACTGACATGG-AGCCTGATGACTATC 1732
|||||
Db 785 AAAGGATTACATGTTATGAAATTTTATGAACTGACATGAAGCCTGATGACTAAT 844

QY 1733 GTGTACTTGGGAAAGTCTCTTGTCTC--ATTGCTGACATGCTTCTGTGTGTCTG 1789
|||||
Db 845 GTGTACTTGGGCAACAGTCTTCTTGTCTTAATTGTGACATGCTTCTGAGGGGG 904

QY 1790 GCCAATGCCAA 1800
|
Db 905 TCTGGGCCCAA 915

RESULT 7
BG740730 817 bp mRNA linear EST 15-MAY-2001
LOCUS 602631591Fl NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4776598 5',
DEFINITION mRNA sequence.
ACCESSION BG740730
VERSION BG740730.1 GI:14051383
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 817)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM10629 row: f column: 23
High quality sequence stop: 794.
location/Qualifiers

FEATURES

Source

1. 817
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4776598"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Skn3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 38.3%; Score 765.8; DB 12; Length 817;
Best Local Similarity 98.8%; Pred. No. 8.5e-109;
Matches 803; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY 1075 CTTCATCTGTTCTAGACAGTACTCACCTCAGCAATTAGCAGGAAGAAATTGAGTGT 1134
|||||
Db 1 CTTCGATCTGTTCTAGACAGTACTCACCTCAGCAATTAGCA-GGAAGAAATTGAGTGT 59

QY 1135 TTTTCTTATGTTCTGTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCT 1194
|||||
Db 60 TTTTCTTATGTTCTGTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCT 119

QY 1195 ACACCGGGGTCTGCTTGATATAAATAACAGCAAGTTATGTGATCTTAAATCAAGCTT 1254
|||||
Db 120 ACACCGGGGTCTGCTTGATATAAATAACAGCAAGTTATGTGATCTTAAATCAAGCTT 179

QY 1255 GATTCAAGAACTGTGTGGCAGCAGATGTCTTCGCTGAATAACATGAAGCTCAGAGAGAC 1314
|||||
Db 180 GATTCAAGAACTGTGTGGCAGCAGATGTCTTCGCTGAATAACATGAAGCTCAGAGAGAC 239

QY 1315 ACCGATCATTTGGTCAACATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAACG 1374
|||||
Db 240 ACCGATCATTTGGTCAACATATTTCCCAAGGTTCAATAGATTCACTCTTTGAAGGAACG 299

QY 1375 TGGTACTTAGTGAAGGTGATGAAGAAGCAGAGAAGAACTTACGCTCGGCGTCCCACTCCA 1434
|||||
Db 300 TGGTACTTAGTGAAGGTGATGAAGAAGCAGAGAAGAACTTACGCTCGGCGTCCCACTCCA 359

QY 1435 AATGATGACACTTTGGATGAAGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCAT 1494
|||||
Db 360 AATGATGACACTTTGGATGAAGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCAT 419

QY 1495 ATTCGAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGAGAAGCTGAAGCA 1554
|||||
Db 420 ATTCGAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGAGAAGCTGAAGCA 479

QY 1555 GCTGTCAATTAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGG 1614
|||||
Db 480 GCTGTCAATTAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGG 539

QY 1615 GTGGGCATGGGGGTGGGGTATGGGAACAAGTTGAGGAATGGGATATCTGGGATTAATTT 1674
|||||
Db 540 GTGGGCATGGGGGTGGGGTATGGGAACAAGTTGAGGAATGGGATATCTGGGATTAATTT 599

QY	1675	AAAGGATTACATGTTATGTATAATTTTTTATGTGACTGCATGGAGCCCTGGATGACTATCGT	1734
Dd	600	AAAGGATTACATGTTATGTATAATTTTTTATGTGACTGCATGGAGCCCTGGATGACTATCGT	659
QY	1735	GTA CTTG GGAAGAAGTCCTCTTGGCTCATATTGTCGACATGCTTCCCTGTTGGTCTGGCCAA	1794
Dd	660	GTACTTGGGAAGAAGTCCTCTTGGCTCATATTGTCGACATGCTTCCCTGTTGGTCTGGCCAA	719
QY	1795	TGCC- AAAT GTA CTG AATGATGTTAAGGGCTCTGTAAAC-TTCATACCCTTTGGCCA	1852
Dd	720	TGCCAAATGTACTCGAATGATGTTAAGGGCTCTGTAAACCTTCATACCCTTTGGCCA	779
QY	1853	TTTGTATGCATGATGTTTGGTTTAAACATGG	1885
Dd	780	TTTGTATGCATGATGTTTAGCGTTTAAACCTTGG	812

RESULT 8
 BG742704
 LOCUS
 DEFINITION
 738 bp mRNA linear EST 15-MAY-2001
 60263291F1 NCI_CGAP_skn3 Homo sapiens cDNA clone IMAGE:4778093 5',
 mRNA sequence.
 ACCESSION
 BG742704
 VERSION
 BG742704.1 GI:14053357
 KEYWORDS
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 738)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.

cdna library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA
 Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: L1A10633 row: e column: 06
 High quality sequence stop: 738.

FEATURES	Location/Qualifiers
source	1..738

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4778093"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP_Skn3"
/note="Organ: skin; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI CGAP Library."

```

ORIGIN

Query Match 36.8%; Score 737; DB 12; Length 738;
Best Local Similarity 100.0%; Pred. No. 2.5e-104;
Matches 737; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1198	CCGGGGTCTGCTCTTGATAAATAACAGCAAGTTATGTGATCTTTAATCAAGCCTGAT	1257
Db	1	CCGGGGTCTGCTCTTGATAAATAACAGCAAGTTATGTGATCTTTAATCAAGCCTGAT	60
QY	1258	TCAAGAACTGGTGTGGCACCAGATGTCCTTGGCGAAACAATGAAGCTCAGAGAGACACC	1317
Db	61	TCAAGAACTGGTGTGGCACCAGATGTCCTTGGCGAAACAATGAAGCTCAGAGAGACACC	120
QY	1318	CATCATTTGGTCAACTATATTTCCCCAGGGTTCAAATAGATTCACTCTTTGAAGGAACGTGG	1377
Db	121	CATCATTTGGTCAACTATATTTCCCCAGGGTTCAAATAGATTCACTCTTTGAAGGAACGTGG	180

QY	1378	TACTTAGTTAGGGTGGATGAAAAACACAGAGAATTACGCTCGGCGTCCCACTCCAAAT	1437
Db	181	TACTTAGTTAGGGTGGATGAAAAACACAGAGAATTACGCTCGGCGTCCCACTCCAAAT	240
QY	1438	GATGACACTTTGGATGAAGAGTAGGACTTTGTGCATTCCAACATAGCACTGAGCATATT	1497
Db	241	GATGACACTTTGGATGAAGAGTAGGACTTTGTGCATTCCAACATAGCACTGAGCATATT	300
QY	1498	CCAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACTGAAGCAGCT	1557
Db	301	CCAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACTGAAGCAGCT	360
QY	1558	GTCAATTAGTAATGGGGAATTAAAGATACTGTGAGGTGCAAGACTTCAGGGTGGGCTG	1617
Db	361	GTCAATTAGTAATGGGGAATTAAAGATACTGTGAGGTGCAAGACTTCAGGGTGGGCTG	420

QY	1618	GGCATGGGGTGGGGGTATGGGAACAGTTGAGGAATGGATATCTGGGGATAATTTTAA	1677
Db	421	GGCATGGGGTGGGGGTATGGGAACAGTTGAGGAATGGATATCTGGGGATAATTTTAA	480
QY	1678	GGATTACATGTTATGTAAATTTTTATGTACTGACATGGAGCCGTGATGACTATCGTGA	1737
Db	481	GGATTACATGTTATGTAAATTTTTATGTACTGACATGGAGCCGTGATGACTATCGTGA	540
QY	1738	CTTGGAAGAAGTCCTCTTGCTCTATTTGCTGCATGCTCTCTGTGTGCTCTGCCAATGC	1797
Db	541	CTTGGAAGAAGTCCTCTTGCTCTATTTGCTGCATGCTCTCTGTGTGCTCTGCCAATGC	600
QY	1798	CAATGTACTGGAATGATGTTAAGGCTCTGTAAAACTTCATACTCTTGGCCATTGT	1857
Db	601	CAATGTACTGGAATGATGTTAAGGCTCTGTAAAACTTCATACTCTTGGCCATTGT	660
QY	1858	ATGCATGATGTTTGGTTTTTAACATGGTATTAATGAATTGTGTACTTCTGTCAAGAAA	1917
Db	661	ATGCATGATGTTTGGTTTTTAACATGGTATTAATGAATTGTGTACTTCTGTCAAGAAA	720
QY	1918	GCAGAGTACTAATCTC	1934
Db	721	GCAGAGTACTAATCTC	737

RESULT 9						
BG683150						
LOCUS	894 bp	mrna	linear	EST 01-MAY-2001		
DEFINITION	602651636F1 NIH_MGC_47 Homo sapiens CDNA clone IMAGE:4761657 5',					
	mrna sequence.					

ACCESSION	BG683150	
VERSION	BG683150.1	GI:139145477
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 894)
COMMENT NIH-MGC <http://mgc.ncl.nih.gov/>.
CONTACT National Institutes of Health, Mammalian Gene Collection (MGC)
UNPUBLISHED (1999)
CONTACT: Robert Strausberg, Ph.D.

Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: LLCM1614 row: h column: 10
High quality sequence stop: 799.

```

FEATURES
  source      location/Qualifiers
              1. .894
              /organism="Homo sapiens"
              /mol_type="mRNA"

```

/db_xref="taxon:9606"
/clone="IMAGE:4761657"
/issue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_47"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

ORIGIN

Query Match	35.3%;	Score 706.2;	DB 12;	Length 894;
Best Local Similarity	92.8%;	Pred. No. 1.3e-99;		
Matches 831;	Conservative 0;	Mismatches 48;	Indels 16;	Gaps 8;
QY	1098	CTCACCTCAGCAATTAGCAGGGAAGAATTGGAGTGTTCCTTATGTTGTTGGC	1157	
Db	2	CTCACCTCAGCAATTAGCAGAGAAGAAGAACTGGAGTGTTCCTTATGTTGTTGGC	61	
QY	1158	TGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACGGGGTCTGCTTGATTA	1217	
Db	62	TGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACGGGGTCTGCTTGATTA	121	
QY	1218	AATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTGGTGGCACC	1277	
Db	122	AATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTGGTGGCACC	181	
QY	1278	AGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGACACCCCATCTTGGTCAACTATAT	1337	
Db	182	AGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGACACCCCATCTTGGTCAACTATAT	241	
QY	1338	TCCCCAGGGTCAATAGATTCACCTTTGAAGGAACGTGTACTTAGTTAGGGTGATGA	1397	
Db	242	TCCCCA-GGTTCAATAGATTCACCTTTGAAGGAACGTGTACTTAGTTAGGGTGATGA	300	
QY	1398	AAAGCAGAGAAGAACTTACGCTCGCGCTCCCACTCCAAATGATGACACTTTGATGAAG	1457	
Db	301	AAAGCAGAGAAGAACTTACGCTCGCGCTCCCACTCCAAATGATGACACTTTGATGAAG	360	
QY	1458	AGTAGAAGCTGTGCATTCAAAACATAGCAACTGAGCATATTCGAAGCCCTGCCAAGAAAT	1517	
Db	361	AGTAGAAGCTGTGCATTCAAAACATAGCAACTGAGCATATTCGAAGCCCTGCCAAGAAAT	420	
QY	1518	ACCAAGACTCCCTGCCACAGCAGAGAACCCTGAAGCAGCTGTCAATTAATAGGGGAACA	1577	
Db	421	ACCAAGACTCCCTGCCACAGCAGAGAACCCTGAAGCAGCTGTCAATTAATAGGGGAACA	480	
QY	1578	TTAAGATACTCTGTGAGTGCAGAACTTCAGGGGTGGGTGGGCATGGGTGGGGTATGG	1637	
Db	481	TTAAGATACTCTGTGAGTGCAGAACTTCAGGGGTGGGTGGGCATGGGTGGGGTATGG	540	
QY	1638	GAACAGTTGGAAGATGGATATCTGGGGAATAATTTTAAAGATTTACATGTTATGAAT	1697	
Db	541	GAACAGTTGGAAGATGGATATCTGGGGAATAATTTTAAAGATTTACATGTTATGAAT	600	
QY	1698	TTTTATGTGACTGACATGAGAGCCGTGATGACTATCGTACTTGGGAAAGTCTC-TTGGC	1756	
Db	601	TTTTATGTGACTGACATGAGAGCCGTGATGACTATCGTACTTGGGAAAGTCTCTTTGGC	660	
QY	1757	TCTATTTGCTGACATGCTTCCTGTGTGTGCTCTGGCCAAATGCCAAATGTAATG	1816	
Db	661	TCTATTTGCTGACATGCTTCCTGTGTGTGCTCTGGCCAAATGCC-AAATGTAATG	719	
QY	1817	TTAAGGCTCTGTAA-ACCTCATAC--TCTTTGGCCATTT--GTATGCATGATGT--T	1869	
Db	720	TTAAGGCTCTGTAAATACTCATACCTCTTTTGGCCATTTGTATTGCCGTATGTCTCT	779	
QY	1870	TGCTTTTAAACATGATATAATGAATGTGTACTTCTGTG-----AGAAGAAAGCAGAG	1923	

Db	780	GGTCGTTTAAACATGCTATCATGAACCTCGTGTACTTCTCGTCAGAAAGAAAGCAGAG	839	
QY	1924	GTAATAATCTCCAAATTAATAATTTTAAACATGTAAAAA	1978	
Db	840	GTAATAATCTCCCAATTTACAAATTTTCTCACCTGGTTCAACAAAAA	894	

RESULT 10
AU127534
LOCUS
DEFINITION
AU127534 NT2RP2 Homo sapiens cDNA clone NT2RP2001505 5', mRNA
sequence.
ACCESSION
AU127534
VERSION
AU127534.1 GI:10952250
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 703)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
CONTACT
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
location/Qualifiers
1..703
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP2001505"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RP2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"

FEATURES		location/Qualifiers
source		1..703
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone="NT2RP2001505"
		/cell_type="teratocarcinoma"
		/cell_line="NT2"
		/clone_id="NT2RP2"
		/note="Vector: pME185Fuj3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"
ORIGIN		
Query Match	33.3%;	Score 666.8; DB 9; Length 703;
Best Local Similarity	97.0%;	Pred. No. 1.7e-93;
Matches 682; Conservative	0;	Mismatches 20; Indels 1; Gaps 1;
QY	823	AAACTGGTTCAGAAATCTCTAGCTCGATGTTGCTGAATGACTTCCTTAATGACCAAGAT 882
Db	1	AAACTGGTTCAGAAATCTCTAGCTCGATGTTGCTGAATGACTTCCTTAATGACCAAGAT 60
QY	883	AGAGATAAAAAATAGTATCTATAGTGGCCTGGAAGCCTTTGGGGAATGTTAATTAGAAGAC 942
Db	61	AGAGATAAAAAATAGTATCTATAGTGGCCTGGAAGCCTTTGGGGAATGTTAATTAGAAANAC 120
QY	943	ACCTACTTTGATAGAGATGTGAGAAGGCATTTATGAAGCCTAGCTCTGAACCTTTCAGT 1002
Db	121	ACCTACTTTGATAGANATGTGANAAGGCATTTATGAAGCCTAGCTCTGAACCTTTCAGT 180
QY	1003	CAGAAAACAAAGGCATCTTTACTTGTATCAAAATCAAAATGGAATATGTACACATCTTCA 1062
Db	181	CAGAAAACAAAGGCATCTTTACTTGTATCAAAATCAAAATGGAATATGTACACATCTTCA 240
QY	1063	GTATATGGTTCCTTGCACTGTCTTACGACAGTACTCACTCAGCAATTAGCAGGGAAG 1122
Db	241	GTATATGGTTCCTTGCACTGTCTTACGACAGTACTCACTCAGCAATTAGCAGGAAN 300

QY	1123	AGAAATTGGAGTGTTTTCTTATGTCTCTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTC	1182
Db	301	ANAAATTGGAGTGTTTTCTTATGTCTCTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTC	360
QY	1183	ACACAAGATGCTACACCGGGTCTGCTCTTGATAAATAACAGCAAGTTATGTGATCTT	1242
Db	361	ACACAAGATGCTACACCGGGTCTGCTCTTGATAAATAACAGCAAGTTATGTGATCTT	420
QY	1243	AAATCAAGGCTTGATTCAAGAACTGGTGTGGCAACAGATGTCCTTCGCTGAAAACATGAAG	1302
Db	421	AAATCAAGGCTTGATTCAAGAACTGGTGTGGCAACAGATGTCCTTCGCTGAAAACATGAAG	480
QY	1303	CTCAGAGAGACACCCATCATTTGGTCAACTATATTTCCCAAGGTTCAATAGATTCACTC	1362
Db	481	CTCAGAAANAGACACCCATCATTTGGTCAACTATATTTCCCAAGGTTCAATATATTCACCTC	540
QY	1363	TTTGAAGGAACGTGGTACTTAACTTAGGGTGGATGAAAAAGCACAGAAAGCTTAACGCTCGG	1422
Db	541	TTTGAAGAAGACGTGGTACTTAACTTAGGGTGGATGAAAAAGCACAGAAAACTTAACGCTCGG	600
QY	1423	CGTCCCACTCCAAATGATGATGACACTTTGGATGAAAGGAGTAGGACTTGTGCATTCAACA-T	1481
Db	601	CGTCCCACTCCAAATGATGACACTTTGGATGAAAGGAGTAGGACTTGTTCATTCAACAATT	660
QY	1482	AGCAACTGAGCATATTTCCAAAGCCCTGCCAAGAAAGTACCAAGA	1524
Db	661	ANCAACTGAACATATTTCCAAANCCCTGCCAANAAAGTACCAANA	703

	RESULT 11			
BX441322				
LOCUS				
DEFINITION	BX441322 Homo sapiens FETAL BRAIN Homo sapiens CDNA clone	1083 bp	mRNA	linear EST 15-MAY-2003
ACCESSION	CSODF016YJ04	5-PRIME,	mRNA sequence.	
VERSION	BX441322			
KEYWORDS	BX441322.1	GI:30789948		
SOURCE	EST.			
ORGANISM	Homo sapiens (human)			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 1083)			
AUTHORS	Li,W.B., Gruber,C., Jesse,J. and Polayes,D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3098.r For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CSODF016DE020D1&cluster=3098.r . Contact :			
	Feng liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscape ID : CSODF016DE02QP1.			

FEATURES	Location/Qualifiers
source	1. .1083

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS06F016YJ04"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo (dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

```

ORIGIN

Query Match	33.2%;	Score 664.2;	DB 13;	Length 1083;
Best Local Similarity	86.2%;	Pred. No. 3.6e-93;		
Matches 799;	Conservative	2;	Mismatches 0;	Indels 126;
				Gaps 1;

QY	134	GCCTTTTACCATTGCCCTGGATCATCTTCTCTTTGAATGCAGAACTTGCTGGCCAAAGATG	193
	157	GCCTTTTACCATTGCCCTGGMTCACTTCTCTTTGAATGCAGAACTTGCTGGCCAAAGATG	216
QY	194	TGGGAATTTGTTGCCCTTGAGATCTATTTCTCTCTCAATATGTTGATCAAGCAGAGTTGG	253
Db	217	TGGGAATTTGTTGCCCTTGAGATCTATTTCTCTCTCAATATGTTGATCAAGCAGAGTTGG	276
QY	254	AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCTGGGCCAGGCCAAGATGG	313
Db	277	AAAAATATGATGGTGTAGATGCTGGAAGTATACCATTTGGCTGGGCCAGGCCAAGATGG	336
QY	314	GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTGCATGACTGTGGTTCAGAACTTTA	373
Db	337	GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTGCATGACTGTGGTTCAGAACTTTA	396
QY	374	TGGAGAGAAATTAACCTTTCTATGATTTGCAATTTGGGCGGCTGGAAAGTTGGAACAGACAA	433
Db	397	TGGAGAGAAATTAACCTTTCTATGATTTGCAATTTGGGCGGCTGGAAAGTTGGAACAGACAA	456
QY	434	TCATGCACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGACTGCGGA	493
Db	457	TCATGCACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGACTGCGGA	516
QY	494	ATACAGATATAGAAGAAATGCACAACTAATGCATGCTATGAGGCAACAGCTGCTGCT	553
Db	517	ATACAGATATAGAAGAAATGCACAACTAATGCATGCTATGAGGCAACAGCTGCTGCT	576
QY	554	TCAATGCTGTTAACTGGAATGAGTCCAGCTCTGGGAT-----	591
Db	577	TCAATGCTGTTAACTGGAATGAGTCCAGCTCTGGGATGAGCGGTATGCCCTGCTAGTTG	636
QY	592	-----	591
Db	637	CAGAGATATTGCTGTATATGCCACAGAAATGCTAGACCTACAGSTGGAGTTGGAGCAG	696
QY	592	-----GGGCTTGTGGGAC	607
Db	697	TAGCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGCTTGTGGGACAC	756
QY	608	ATATGCACATGCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG	667
Db	757	ATATGCACATGCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG	816
QY	668	ATGGAATCTCTCCATACAGTGTCTAAGTGTGATTTAGACCGCTGCTACTCTGTCTACT	727
Db	817	ATGGAATCTCTCCATACAGTGTCTAAGTGTGATTTAGACCGCTGCTACTCTGTCTACT	876
QY	728	GCAAAAAGATCCATGCCCGCAGTGGCAGAAAGAGGGAATGATAAAGATTTTACCTTGAATG	787
Db	877	GCAAAAAGATCCATGCCCGCAGTGGCAGAAAGAGGGAATGATAAAGATTTTACCTTGAATG	936
QY	788	ATTTTGGCTTCATGATCTTTGCACTCACCATATTGTAACTGGTTCAGAAATCTCTAGCTC	847
Db	937	ATTTTGGCTTCATGATCTTTGCACTCACCATATTGTAACTGGTTCAGAAATCTCTAGCTC	996
QY	848	GGATGTTGCTGAATGACTTCTTAAATGACCAGAAATAGAGATAAATATGATATCTATAGTG	907
Db	997	GGATGTTGCTGAATGACTTCTTAAATGACCAGAAATAGAGATAAATATGATATCTATAGTG	1056
QY	908	GCCTGGAAGCCTTTGGGAGTGTAAAT	934
Db	1057	GCCTGGAAGCCTTTGGGAGTGTAAAT	1083

RESULT 12			
BG683791/c			
LOCUS	BG683791	795 bp	mRNA linear EST 01-MAY-2001

DEFINITION 602651636T1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:4761657 3',
mRNA sequence.
ACCESSION BG683791
VERSION BG683791.1 GI:13915188
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 795)
NIH-MGC http://mgc.nci.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
plate: L1CM1614 row: h column: 10
High quality sequence start: 24
High quality sequence stop: 758.
location/Qualifiers
1. 795

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4761657"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_47"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 32.8%; Score 657.6; DB 12; Length 795;
Best Local Similarity 95.0%; Pred. No. 4.3e-92;
Matches 756; Conservative 0; Mismatches 29; Indels 11; Gaps 7;

QY 1117 GCGAAGAAATTGGAGTGTTCCTATAGTTCGTTGGCTGCCACTCTGTACTCTCT 1176
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 794 GGAACGAGAACCGGAGTGTTCCT--AGGTTCTGGCTGGCTGCCACTGTACTCTCGT 737
QY 1177 AAAGTCACACAGATGCTACACCGGGTCTGCTCTTGAT-AAATAACAGCAAGTT--T 1232
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 736 AAAGTCACACAGACTGCTACACCGGGCTGCTCTTGATCCAAATAACAGCAAGTTTAT 677
QY 1233 ATGTGATCTTAATCAAGGCTTGATTCAA--GAAGTGGTGGGACCCAGATGTTGCGCT 1290
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 676 GTGACTCTTAACTCAAGGCTTGATTCAAAGAACTGGTGTGGACCCAGATGTTGCGCT 617
QY 1291 GAAAAATGAAGCTCAGAGAGACCCCATCTTGGTCAACTATATTCCCGAGGGTTCA 1350
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 616 GAACACATGAAGCTCAGAGAGACCCCATCTTGGTCAACTATATTCCCGAGGGTTCA 557
QY 1351 ATAGATTCACCTCTTGAAGGAAGCGGTAC-TTAGTTAGGGTGATGAAAAACACAGAG 1409
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 556 ATAGATTCACCTCTTGAAGGAAGCGGTACTTTAGTTAGGGTGATGAAAAACACAGAG 497
QY 1410 AACTTACGCTGGGCTCCCACTCCAATGATGACACTTTGGATGAAGAGTAGACTTGT 1469
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 496 AACTGACGCTGGGCTCCCACTCCAATGATGACACTTTGGATGAAGAGTAGACTTGT 437

QY 1470 GCATTCAACATAGCAACTGAGCATATTCCAAAGCCCTGCCAAGAAAGTACCAAGACTCCC 1529
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 436 GCATTCAACATAGCAACTGAGCATATTCCAAAGCCCTGCCAAGAAAGTACCAAGACTCCC 377
QY 1530 TGCCACAGCAGCAGAACCTGAGCAGCTGTCTATTAGTAATGGGAACATTAGACTACT 1589
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 376 TGCCACAGCAGCAGAACCTGAGCAGCTGTCTATTAGTAATGGGAACATTAGACTACT 317
QY 1590 GTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGTGGGTATGGAAACAGTTGGAG 1649
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 GTGAGGTGCAAGACTTCAGGGTGGGGTGGGCATGGGGTGGGTATGGAAACAGTTGGAG 257
QY 1650 GAATGGGATATCTGGGGATTAATTTTAAAGGATTACATGTTATGTAATTTTATGTGACT 1709
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 GAATGGGATATCTGGGGATTAATTTTAAAGGATTACATGTTATGTAATTTTATGTGACT 197
QY 1710 GACATGAGCCCTGGATGACTATCGTGTACTTGGGAAAGTCTTTGCTCTATTGCTGAC 1769
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 196 GACATGAGCCCTGGATGACTATCGTGTACTTGGGAAAGTCTTTGCTCTATTGCTGAC 137
QY 1770 ATGCTTCCTGTTGTGCTGCGCCAAATGCCAAATGTAATGATGTTAAGGGC-TCGTG 1828
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 136 ATGCTTCCTGTTGTGCTGCGCCAAATGCCAAATGTAATGATGTTAAGGGCTTCGTG 77
QY 1829 TAAACTTCATACCTCTTTGGCCATTGTGTATGCATGATGTTGTTTTTAAACATGGTA- 1887
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76 TAAACTTCATACCTCTTTGGCCATTGTGTATGCATGATGTTGTTTTTAAACATGGAAC 17
QY 1888 TAATGAATGTGTACT 1903
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16 TACTGAATGTGTACT 1

RESULT 13
BG698557 733 bp mRNA linear EST 07-MAY-2001
LOCUS 602658594F2 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4801702 5',
DEFINITION mRNA sequence.
ACCESSION BG698557
VERSION BG698557.1 GI:13965967
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 733)
NIH-MGC http://mgc.nci.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
plate: L1AM10694 row: 1 column: 23
High quality sequence start: 33
High quality sequence stop: 731.
location/Qualifiers
1. 733

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4801702"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP_Skn3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 32.2%; Score 644.2; DB 12; Length 733;
Best Local Similarity 98.3%; Pred. No. 5.2e-90;
Matches 683; Conservative 0; Mismatches 8; Indels 4; Gaps 3;

QY 1198 CCGGGGTGCTCTTGATMAAAATA CAGCAAGTTTATGTGATCTTAAATCAAGGCTTGAT 1257
DB 40 CCGCGGTGCTCTTGATMAAAATA CAGCAAGTTTATGTGATCTTAAATCAAGGCTTGAT 99
QY 1258 TCAAGAACTGGTGTGGCAC CAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGACACC 1317
DB 100 TCAAGAACTGGTGTGGCAC CAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGACACC 159
QY 1318 CATCATTTGGTCAACTATATTTCCCAAGGGTTCAATAGATTCACTCTTTGAAGAAAGTGG 1377
DB 160 CATCATTTGGTCAACTATATTTCCCAAGGGTTCAATAGATTCACTCTTTGAAGAAAGTGG 219
QY 1378 TACTTAGTAGGGTGATGAAAAGCA CAGAAGACTTACGCTCGGCGTCCCACTCCAAAT 1437
DB 220 TACTTAGTAGGGTGATGAAAAGCA CAGAAGACTTACGCTCGGCGTCCCACTCCAAAT 279
QY 1438 GATGACACTTTGGATGAAGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATT 1497
DB 280 GATGACACTTTGGATGAAGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATT 339
QY 1498 CCAAGCCCTGCCAAGAAAGTACC AAGACTCCCTGCCACAGCAGACAGAACTGAAGCAGCT 1557
DB 340 CCAAGCCCTGCCAAGAAAGTACC AAGACTCCCTGCCACAGCAGACAGAACTGAAGCAGCT 399
QY 1558 GTCATTTAGTAATGGGAA CATTAAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTG 1617
DB 400 GTCATTTAGTAATGGGAA CATTAAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGTG 459
QY 1618 GGCATGGGGTG- GGGGTATGGGAA CAGTTGGAGGAATGGGATATCTGGGGATAATTTAA 1676
DB 460 GGCATGGGGTGCGGGGTATGGGAA CAGTTGGAGGAATGGGATATCTGGGGATAATTTAA 519
QY 1677 A-GGATTAATGTATGTAATTTTAAATTTTATGTGACTGACATGAGCCCTGATGACTATG 1735
DB 520 ACGGATTAATGTATGTAATTTTAAATTTTATGTGACTGACATGAGCCCTGATGACTATG 579
QY 1736 TACTTGGGAAAGTCTCTTGGCTCTATTTGCTGACATGCTTCTGTGSGTCTGGCCAAT 1795
DB 580 TACTTGGGAAAGTCTCTTGGCTCTATTTGCTGACATGCTTCTGTGSGTCTGGCCAAT 639
QY 1796 GCCAAATGTACTCGAATGATGTTAAGGGCTGTGAAAACCTCATACCTTTGGCCATTT 1855
DB 640 GCCAAATGTACTCGAATGATGTTAAGGGCTGTGAAAACCTCATACCTTTGGCCATTT 699
QY 1856 GTATGCAATGATGTTGTTTTTAAACATGATATA 1890
DB 700 GTATGCAATGATG--TCGGTTTTAAACATGATATA 732

RESULT 14

CA512392 789 bp mRNA linear EST 15-NOV-2002
LOCUS UI-R-FJ0-cpz-j-19-0-UI.r1 UI-R-FJ0 Rattus norvegicus cDNA clone
DEFINITION UI-R-FJ0-cpz-j-19-0-UI 5', mRNA sequence.

ACCESSION CA512392
VERSION CA512392.1 GI:25003346
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 789)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE

PUBMED

97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).

FEATURES

source

1. 789
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-FJ0-cpz-j-19-0-UI"
/tissue_type="embryo"
/dev_stage="embryo"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-FJ0"
/note="vector: pYX-asc; Site_1: EcoR I; Site_2: Not I;
UI-R-FJ0 is a cDNA library containing the following
tissue(s): rat embryo. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dr)18 tail. The
sequence tag for this library is CATCTCTACT. This library
was created for the University of Iowa Program for Rat
Gene Discovery and Mapping (Val Sheffield, Bento Soares
and Tom Casavant)"

ORIGIN

Query Match 32.1%; Score 643.6; DB 14; Length 789;
Best Local Similarity 88.3%; Pred. No. 6.3e-90;
Matches 697; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 710 GCTGCTACTCTGTCTACTGCAAAAAGATCCATGCCAGTGGCAGAAAGAGGAATGATA 769
DB 1 GCTGCTATTTCTGTCTACCGCAAAAAGATCCGGGCCACAGTGGCAGAAAGAGGATA 60
QY 770 AAGATTTTACCTTGATGATTTTGGCTTCATGATCTTCACTCAACCATATTGTAACCTGG 829
DB 61 AAGATTTTACCTTGATGATTTTGGCTTCATGATCTTCACTCAACCATATTGTAACCTGG 120
QY 830 TTCAGAAATCTAGCTCGATGTTGCTGAATGACTTCCCTTAATGACCAGAAATAGAGATA 889
DB 121 TGCAGAAATCTTAGCTAGATGTTCTCTGAATGACTTCTTAAAGATCAAAAACAGAGACA 180
QY 890 AAAATAGTATCTATAGTGGCCCTGGAAGCCCTTGGGATGTTAAATTGAAGACACTTACT 949
DB 181 AAAACAGTATTTACAGTGGGCTGGAAGCCCTTGGGATGTTAAATTGAAGACTTACT 240
QY 950 TTGATAGAGATGTGAGAGAGGCAATTATGAAGGCTAGCTGTGAACCTTCAAGTACAGAAA 1009
DB 241 TCGACAGAGATGTGAAAAGGCAATTATGAAGGCTAGCTGTGAGCTATTCAACAGAAAA 300
QY 1010 CAAAGGCATCTTTACTTGTATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATG 1069
DB 301 CAAAGGCATCTTTGCTTGTATCGAATCAAAATGGAACATGTACACATCTCTGTATACG 360
QY 1070 GTTCCCTTGCACTGTCTTAGACACAGTACTCCTCAGCAATTAGCAGGAAGAGAAATTG 1129

Db 361 GTTCCCTTGCTTCTGTCTTGCCACAGTACTCACCTCAACAGTTGGCCGGAGAGGATTG 420

QY 1130 GAGTGTTTCTTATGGTTCTGGTTTGGCTGCCACTCTGTACTCTTTAAAGTCACACAAG 1189

Db 421 GAGTGTCTCTTACGGTCTTGCTTGCTGCCACACCTTACTCCCTTAAAGTCACACAAG 480

QY 1190 ATGCTACACCGGGGCTGCTCTTGATTAATAACAGCAAGTTATGTGATCTTAAATCAA 1249

Db 481 ATGCCACACACGAGATCTGCTCTTGACAAATAACAGCAAGTTATGTGACCTTAAAGTCAA 540

QY 1250 GGCTTGATTCAAGAACTGCTGTGTGGCACCCAGATGTCTTCGCTGAATAACATGAGCTCAGAG 1309

Db 541 GGCTTGACTCAAGAACGTGTGTGGCACCCAGACGCTTTGTGTGAATAACATGAAAGCTCAGAG 600

QY 1310 AGGACACCCCATCATTTGGTCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAG 1369

Db 601 AGGACACACATCATCTTAGCCAACTATATTGCCAGTGTTCATATGATTCACTCTTCGAAG 660

QY 1370 GAACGTGTAAGTTAGTGAAGGTGGATGAAAAGCACAGAGAAGCTTACGCTCGGCTGCCA 1429

Db 661 GAACATGTATCTAGTCAAGTGGATGAAAAGCACAGAGAAGCTTACGCCCGCGTCCCT 720

QY 1430 CTCCAATGATGACACTTTGGATGAAGAGTAGGACTTGTGCAATTCAAACATAGCAACTG 1489

Db 721 CCACANNATGACACAGTTTGGATGAAGGAGTGGGACTTGTGCAATTCANACACAGCTACAG 780

QY 1490 AGCATATTTC 1498

Db 781 AGCATATTTC 789

RESULT 15	
LOCUS	BX419944
DEFINITION	BX419944 1201 bp mRNA linear EST 13-MAY-2003
ACCESSION	BX419944 Homo sapiens FETAL BRAIN Homo sapiens CDNA clone
VERSION	CS0DF022YE18 5-PRIME, mRNA sequence.
KEYWORDS	BX419944
SOURCE	BX419944.1 GI:30650862
ORGANISM	EST.
REFERENCE	Homo sapiens (human)
AUTHORS	Homo sapiens
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL	1 (bases 1 to 1201)
COMMENT	Li, W.B., Gruber, C., Jessee, J. and Polayes, D. Full-length cDNA libraries and normalization Unpublished (2001)
	Contact: Genoscope
	Genoscope - Centre National de Sequencage
	BP 191 91006 EVRY cedex - France
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
	library was constructed by life Technologies, a division of
	Invitrogen. This sequence belongs to sequence cluster 3098.r For
	more information about this cluster, see
	http://www.genoscope.cns.fr/
	cgi-bin/cluster.cgi?seq=CS0DF022BC09QP1&cluster=3098.r. Contact :
	Feng Liang Email : fliang@lifetech.com URL :
	http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
	Paradise Avenue Genoscope sequence ID : CS0DF022BC09QP1.
FEATURES	Location/Qualifiers
Source	1..1201

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022YE18"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: brain; Vector: pCMWSPORT 6; 1st strand cDNA
was primed with a NotI-oligo (dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMWSPORT 6

```

vector. Library was not normalized."

Query Match	31.3%	Score 626;	DB 13;	Length 1201;
Best Local Similarity	84.0%;	Pred. No. 2.7e-87;		
Matches 818; Conservative	4;	Mismatches 19;	Indels 133;	Gaps 5;

QY	100	ACTCCCTTTCCTCTGCTGCCGCTCGGTCACGCGCTGCTCTTTTCACCATGCTGGATCACTT	159
Db	45	ATTCCCTTTCCTCTGCTGCCGCTCGGTCACGCGCTGCTCTTTTCACCATGCTGGATCACTT	104
QY	160	CCTTTGAATGCAGAAGCTTGCTGCGCCAAAAGATGTGGGAATGTGCCCTTGAGATCTAT	219
Db	105	CCTTTGAATGCAGAAGCTTGCTGCGCCAAAAGATGTGGGAATGTGTGCCCTTGAGATCTAT	164
QY	220	TTTCCTTCTCAATATGTGTGATCAACAGAGAGTTGGAAAAATATGATGGTTAGATGCTGGA	279
Db	165	TTTCCTTCTCAATATGTGTGATCAACAGAGAGTTGGAAAAATATGATGGTTAGATGCTGGA	224
QY	280	AAGTATACCATTTGGCTTGGGCCAGGCCAAGATGGGCTTCTGCACAGATAGAGAAGATATT	339
Db	225	AAGTATACCATTTGGCTTGGGCCAGGCCAAGATGGGCTTCTGCACAGATAGAGAAGATATT	284
QY	340	AACTCTCTTTGCATGACTGTGGTTCAAGATCTTATGAGAGAAATACTTTCCTATGAT	399
Db	285	AACTCTCTTTGCATGACTGTGGTTCAAGATCTTATGAGAGAAATACTTTCCTATGAT	344
QY	400	TGCATTGGGCGGCTGGAAGTTGGAACAGAGACAATCATCGACAAATCAAAAGTCTGTGAAG	459
Db	345	TGCATTGGGCGGCTGGAAGTTGGAACAGAGACAATCATCGACAAATCAAAAGTCTGTGAAG	404
QY	460	ACTAATTTGATGCAGCTGTGTTGAAGAGTCTGGGAATACAGATATAGAAAGAAATCGACACA	519
Db	405	ACTAATTTGATGCAGCTGTGTTGAAGAGTCTGGGAATACAGATATAGAAAGAAATCGACACA	464
QY	520	ACTAATGCATGCTATGAGGACACAGCTGCTGTCTTCAATGCTGTAACTGGATGAGTCC	579
Db	465	ACTAATGCATGCTATGAGGACACAGCTGCTGTCTTCAATGCTGTAACTGGATGAGTCC	524
QY	580	AGCTCTTGGGAT-----	591
Db	525	AGCTCTTGGGATGACGGTATGCCCTGCTAGTTGCAGAGATATTGCTGATATATGCCACA	584
QY	592	-----	591
Db	585	GGAATGCTAGACCTACAGGTGAGCTTGGAGCAGTAGCTCTGCTAATTGGCCCAATGCT	644
QY	592	-----GGGCTTTCGTGGGACACATATATGCAACATGCTTATTTTAC	633
Db	645	CCTTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACATGCTTATTTTAC	704
QY	634	AAGCTGATATGCTATCTGAATATCCTATAGTAGATGAAAACTCTCCATACAGTGCTAC	693
Db	705	AAGCTGATATGCTATCTGAATATCCTATAGTAGATGAAAACTCTCCATACAGTGCTAC	764
QY	694	CTCAGTCATTAGACCGCTGCTACTCTGTCTACTGCAAAAAAGATCCATGCCAGTGGCAG	753
Db	765	CTCAGTCATTAGACCGCTGCTACTCTGTCTACTGCAAAAAAGATCCATGCCAGTGGCAG	824
QY	754	AAAGAGGGAATGATAAGA-TTTTACCTTGAATGATTTTGGCTTCATGATCTTTCACTC	812
Db	825	AAAGAGGGAATGATAAGATTTTTCCTTGAATGATTTTGGCTTCATGATCTTTCACTC	884
QY	813	ACCATATTGTAACTGGTTCAGAAATCTCTAGCTCGAGTGTGCTGAATGACTTCTCTAA	872
Db	885	ACCATATTGTAACTGGTTCAGAAATCTCTAGCTCGAGTGTGCTGAATGAC-TCCTTAA	943
QY	873	TGACCAGAATAGAGATAAAAAATGATCTATAGTGGCCTGGAAGCTTTGGGAGTGTAA	932
Db	944	TGMCAGATAGAGATAAAATAGTA-----CTATAGTGGCCTGG-AGCCTTGGGGAAGTAA	998
QY	933	ATTAAGAAGACACCT	946

Fri Jun 25 07:33:52 2004

us-10-622-516-1.rst

Page 16

Db 999 TWAGARGCCTCT 1012

Search completed: June 24, 2004, 11:14:15
Job time : 4695 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 24, 2004, 13:22:38 ; Search time 4601 Seconds
(without alignments)
4502.928 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPGLPLNAEACWPKDVGIV.....PRLPATAEPEAAVISNGEH 478

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 segs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool_p/US10622516/runat_23062004_162633_648/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10622516@CGN_1_1_5600@runat_23062004_162633_648 -NCPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrc:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2511	100.0	2002	6	AR399482	AR399482 Sequence
2	2480	98.8	1563	9	BT007302	BT007302 Homo sapi
3	2480	98.8	1650	9	HUM3H3M	L25798 Homo sapien
4	2480	98.8	2068	9	BC000297	BC000297 Homo sapi
5	2439	97.1	1685	9	HSCOAS	X66435 H.sapiens m
6	2403.5	95.7	3195	9	AK095492	AK095492 Homo sapi
7	2362	94.1	3275	6	AX700129	AX700129 Sequence
8	2362	94.1	3275	6	AX827303	AX827303 Sequence
9	2362	94.1	3275	10	RNCHMCA	X52625 Rattus norv
10	2348	93.5	2703	10	BC023851	BC023851 Mus muscu
11	2348	93.5	3202	10	BC034317	BC034317 Mus muscu
12	2348	93.5	3260	10	BC029693	BC029693 Mus muscu
13	2344	93.3	3278	10	BC031363	BC031363 Mus muscu
14	2097	83.5	1874	5	CHKMGCOAS	M60657 Chicken HMG
15	2089	83.2	1824	6	165510	I65510 Sequence 1
16	2003	79.8	3098	5	BC042929	BC042929 Xenopus 1
17	1957.5	78.0	3205	5	BC049456	BC049456 Danio rer
18	1691	67.3	127832	9	AL356361	AL356361 Human DNA
19	1691	67.3	170834	2	AC021792	AC021792 Homo sapi
20	1513	60.3	2058	6	AX337379	AX337379 Sequence
21	1513	60.3	2058	6	AX411166	AX411166 Sequence
22	1513	60.3	2058	9	HSBMGCOAS	X83618 H.sapiens m
23	1513	60.3	2098	9	BC044217	BC044217 Homo sapi
24	1511.5	60.2	1994	6	AX163779	AX163779 Sequence
25	1511.5	60.2	1994	10	RATHMGOA	M33648 Rat mitocho
26	1494	59.5	1991	10	AK098104	AK098104 Mus muscu
27	1493	59.5	1994	10	BC024744	BC024744 Mus muscu
28	1493	59.5	2011	10	BC014714	BC014714 Mus muscu
29	1466	58.4	1744	4	SSU90884	U90884 Sus scrofa
30	1396	55.6	163423	2	AC123738	AC123738 Mus muscu
31	1365.5	54.4	144405	9	AC114947	AC114947 Homo sapi
32	1362.5	54.3	142657	2	AC008773	AC008773 Homo sapi
33	1359	54.1	1658	3	BGHMG	X73679 B.germanica
34	1353	53.9	1296	10	MMU12790	U12790 Mus musculu
35	1337	53.2	2588	3	AY051743	AY051743 Drosophil
36	1337	53.2	97346	2	AC017201	AC017201 Drosophil
37	1337	53.2	157835	3	AC009356	AC009356 Drosophil
38	1337	53.2	159640	3	AC008230	AC008230 Drosophil
39	1337	53.0	236502	3	AE003807	AE003807 Drosophil
40	1332	53.0	236502	2	AC098186	AC098186 Rattus no
41	1272.5	50.7	1716	3	BGHMGCOA3	X77516 B.germanica
42	1223.5	48.7	1881	3	AK112832	AK112832 Ciona int
43	1204.5	48.0	259650	2	AC107176	AC107176 Rattus no
44	1179	47.0	2015	3	AF166002	AF166002 Dendrocto
45	1115	44.4	1542	9	HSU12789	U12789 Human clone

RESULT 1

ALIGNMENTS

AR399482
LOCUS AR399482 2002 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 1 from patent US 6620608.
ACCESSION AR399482
VERSION AR399482.1 GI:40141523
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2002)
AUTHORS Gong, F., Yan, C., Di Francesco, V. and Beasley, E.M.
TITLE Isolated human synthase proteins
JOURNAL Patent: US 6620608-A 1 16-SEP-2003;
FEATURES
location/Qualifiers
1..2002
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 4.42e-205 Length: 2002
Score: 2511.00 Matches: 478
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-622-516-2 (1-478) x AR399482 (1-2002)

QY 1 MetProGlySerIeuProLeuAenAlaGluAlaCysTrpProIysAspValGlyIleVal 20
Db 145 ATGCCTGATCACTTCTTGAAATGCAGAGCTTGCCCAAAAGATGTGGAAATTGTT 204
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluIlySTyrAsp 40
Db 205 GCCCTTGAGATCTATTTCTCTTCAATATGTGTGATCAAGCAGAGCTTGAAAAATATGAT 264
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaIlyMetGlyPheCysThr 60
Db 265 GGTGTAGATGCTGAAAGATATACCATTTGGCTGGCCAGCCAAAGATGGCTTCTGCACA 324
QY 61 AspArgGluAspIleAsnSerIeuCysMetThrValAlaGlnAsnLeuMetGluArgAsn 80
Db 325 GATAGAGAAGATATTAACCTCTCTTTCATGACTGTGTTCAGAACTTATGGAGAGAAAT 384
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 385 AACCTTTCCTATGATTGCATTGGCGCGCTGGAGATTGGAAAGACAGACAATCATGCACAAA 444
QY 101 SerLysSerValIlySThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 445 TCAAAGTCTGTGAAGACTAATTGTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA 504
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 505 GAAGGAATCGACACACTAATGCATGCTATGAGAGGCACACGCTGCTCTCAATGCTGTT 564
QY 141 AsnTripleGluSerSerSerTrpAspGlyLeuArgGlyThrHisMetGlnHisAlaTyr 160
Db 565 AACTGATGAGTCCAGCTCTGGGATGGGCTTCGTGGACACATATGCACATGCTCAT 624
QY 161 AspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIle 180
Db 625 GATTTTACACAGCTGATATGCTATCTGAATATCCTATAGTAGATGGAACACTCTCCATA 684
QY 181 GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAla 200
Db 685 CAGTGTACTCTCAGTGCATTAAGACCGCTGCTACTCTGCTACTGCAAAAAGATCCATGCC 744
QY 201 GlnTyrGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIle 220
Db 745 CAGTGGCAGAAAGAGGGAATGATTAAGATTTTACCTGAATGATTTTGCTTCATGATC 804

QY 221 PheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp 240
Db 805 TTTCACCTCACCATATTTGTAACTGGTTCAAGAAATCTTAGCTCGGATGTTGCTGAATGAC 864
QY 241 PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGly 260
Db 865 TTCCTTAATGACCAAGATAGAGTAAATAATAGTATCTATAGTGGCTGGAAAGCCCTTGGG 924
QY 261 AspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAla 280
Db 925 GATGTTAAATTAGAAGACACCTACTTGTAGATGTGAGAGGCATTTATGAAGCT 984
QY 281 SerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGly 300
Db 985 AGCTCTGAACCTTTCAGTCAGAAACAAAGGCATCTTACTTGTATCAATCAAAATGGA 1044
QY 301 AsnMetTyrThrSerSerValTyrGlySerIleuAlaSerValIleuAlaGlnTyrSerPro 320
Db 1045 AATATGTACACATCTTCAGTATATAGTTCCCTTGCAATCTGTTCTAGCACAGTACTCACT 1104
QY 321 GlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThr 340
Db 1105 CAGCAATTAGCAGGGAAGAGAAATTGAGTGTCTTATAGTCTGTGCTGGCCGCACT 1164
QY 341 LeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThr 360
Db 1165 CTGTACTCTCTTAAAGTCACACAAAGATGTCACCGGGTCTGCTCTTGATAAATTAACA 1224
QY 361 AlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPheAspVal 380
Db 1225 GCAAGTTATGTGATCTTAAATCAAGCCTTGATTCAGAAGACTGTGTGGCACCAATGTC 1284
QY 381 PheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIleProGln 400
Db 1285 TTCGCTGAAGACATGAAGCTCAGAGAGACACCCATCATTTGTCACACTATATTTCCCCAG 1344
QY 401 GlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGluLysHis 420
Db 1345 GGTTCATATGATTCACTCTTGAAGGAACGTGTACTTAGTAAAGGTGATGAAGAAAGCAC 1404
QY 421 ArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGlyValGly 440
Db 1405 AGAAGAACTTAACGCTCGGCGTCCCACTCCAATGATGACACTTTGGATGAAGAGTAGGA 1464
QY 441 LeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysValProArg 460
Db 1465 CTGTGCATTCAAAACATAGCAACTGAGCAGTAATTCGAAGCCCTGCCAAGAAAGTACCAAGA 1524
QY 461 LeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db 1525 CTCCTGCCACAGCAGCAAGAACCTGAAGCAGCTGTCAATTAGTAAATGGGGAACAT 1578
RESULT 2
BT007302 1563 bp mRNA linear PRI 13-MAY-2003
LOCUS BT007302
DEFINITION Homo sapiens 3-hydroxy-3-methylglutaryl-coenzyme A synthase 1
(soluble) mRNA, complete cds.
ACCESSION BT007302
VERSION BT007302.1 GI:30583442
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1563)
AUTHORS Kalinine, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S.,
Koundinya, M., Raphael, J., Moreira, D., Kelley, T., Labaer, J., Lin, Y.,
Phelan, M. and Farmer, A.
TITLE Cloning of human full-length cDNAs in BD Creator(TM) system Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1563)
AUTHORS Kalinine, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S.,

TITLE Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., PheJan,M. and Farmer,A.
JOURNAL Direct Submission
COMMENT Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the Sali and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after Sali site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.

location/Qualifiers

FEATURES

source

CDS

1.1563
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GH00878X1.0"
/clone_lib="BD Creator(TM) CDS library derived from MGC collection"
/lab_host="DH5alpha T1 resistant"
/note="Vector: pDNR-Dual"
1.1563
/codon_start=1
/product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)"
/protein_id="AAP35966.1"
/db_xref="GI:30583443"
/translation="MGSLPLNAEACWPKDVGIVALEIYFSPQYVDOAELEKYGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVVQNLMEERNLSYDCIGRLVETETIIDSKS SVKTNLMQLFEESGNTDIEGIDTTNACYCGTAAVFNANWMISSSWDGRVALVAGDI AVYATGNARPTGGVAVALLIGPNAPLIFERGLRGTMOHAYDFYKPDMLSEYPIDG KLSIQCYLSALDRCSVYCKKIHAQWQKEGNDKDFTLNDFGFIHSPYCKLVQKSLA RMLNDFLNDQNRDKNSIYSGLEAFGDKLEDYFDRDVEKAFMKASSELSQKTKAS LTVSNQNGNMVTSIVYGSGLASVLAQYSPOQLAGKRIGVFSYSGLAATLYSLKVTQDA TPGSALDKITASLCDKSRIDSRRTGVA PDVPAENMKLRDTHLTVNYIPQSIDSLFE GTWYLVVRVDEKHKRTYARRPNPNDTLDEGVGLVHSNIATEHIIPSPAKKYRRLPATAA EPEAAVISNGEH"

ORIGIN

Alignment Scores:

Pred. No.: 1.42e-202 length: 1563
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 9 Gaps: 1

US-10-622-516-2 (1-478) x BT007302 (1-1563)

QY 1 MetPProGlySerLeuProLeuAnaIaGluAlaCysTRProLyAspValGlyIleVal 20
Db 1 ATGCCTGCATCACTTCCTTTGAATGCAGAAAGCTTGCTGGCCAAAGATGTGGGAATTGTT 60
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnaIaGluLeuGluLySTyrAsp 40
Db 61 GCCCTTGAGATCTATTTCCCTTCATATATGTTGATCAAGCAGAGATTGGAAAAATATGAT 120
QY 41 GlyValaAspAlaGlyLySTyrThrIleGlyLeuGlyGlnaIaIaIaMetGlyPheCysThr 60
Db 121 GGTTGATGCTGGAAGTATACATTGGCTTGCGCCAGGCCAAGATGGGCTTCTGCACA 180
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 181 GATAGAGAAGATATTAACTCTCTTGATGACTGTGTTCAAGATCTTATGAGAGAAAT 240
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 241 AACCTTCTATGATTCATTGGCGCGCTGGAAGTTGGAACAGACACAATCATCGACAAA 300

QY 101 SerLySerValLySThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 301 TCAAAAGTCTGGAAGACTAATTTGATGCAGCTGTTGAAGAGTCTGGAAATACAGATATA 360
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 361 GAAGGAATGCACACAACATAATGCATGCTATGAGGACACAGCTGCTCTCAATGCTGTT 420
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 421 AACTGATTGAGTCCAGCTCTGGGATGGACGGTATGCCCTGCTAGTTCAGAGAGATATT 480
QY 149 ----- 149
Db 481 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGATTGGACAGTAGCTTGCTA 540
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 541 ATTGGGCCAAATGCTCCTTTAATTTTGAAACGAGGGCTTCGTGGACACATATGCAACAT 600
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 601 GCCATGATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGATGAAAACTC 660
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIle 198
Db 661 TCCATACAGTGTACTCCTCAGTGCATTAGACCCGCTGCTACTGCTACTGCCAAAAGATC 720
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 721 CATGCCCAGTGGCAGAAAGAGGAATGATTAAGATTTAACCTTGAATGATTTTGGCTTC 780
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeu 238
Db 781 ATGATCTTCACTCACCAATATTGTAACCTGGTTCAGAAATCTCTAGCTCGGATGTTGCTG 840
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 841 AATGACTTCTTAATGACCAAGAATAAGATAAAAATATATCTATAGTGCGCTGGAAGCC 900
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 901 TTTGGGAGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGAAGCATTTATG 960
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 961 AAGGCTAGCTCTGAACCTTTCAGTCAAGAAAACAAGGCATCTTACTGTATCAATCAA 1020
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1021 AATGGAATATATGACACATCTTCAGTATATGTTCCCTTGCACTGTTCTAGCACAGTAC 1080
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValaPheSerTyrGlySerGlyLeuAla 338
Db 1081 TCACCTCAGCAATTAGCAGGGAAGAAATTGAGTGTTTTCTTATGTTCTGTGTTGGCT 1140
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1141 GCCACTCTGTACTCTCTTAAAGTCACACAGATGCTACACCGGGGCTGCTTGATATAA 1200
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValaIaPro 378
Db 1201 ATAAACAAGATTATGTGATCTTAAATCAAGGCTTGATTTCAAGAACTGGTGGCACCA 1260
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValaAsnTyrIle 398
Db 1261 GATGCTTCGCTGAAGACATGAAAGCTCAGAGAGGACACCATCATTTGTTCAACTATATT 1320
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValaAspGlu 418
Db 1321 CCCAGGGTTCAATAGATTCACTCTTGAAAGGAACGTGTACTTAGTGGGTGATGAA 1380
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438

|||||
Db 1381 AAGCACAGAAAGACTTACGCTCGCGCTCCACATCCAAATGATGACACTTTGGATGAAGGA 1440
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaIleValVal 458
Db 1441 GTAGGACTTGTCATTCAACACATGCACTGAGCATATTCCAAAGCCCTGCCAAGAAAGTA 1500
QY 459 ProArgLeuProAlaThrAlaIleGluProGluIleAlaValIleSerAsnGlyGluHis 478
Db 1501 CCAAGACTCCCTGCGCACAGCAGCAAACTGAAGCAGCTGTCAATTAGTAATGGGGAACAT 1560
RESULT 3
HUM3H3M 1650 bp mRNA linear PRI 24-FEB-1995
LOCUS Homo sapiens 3-hydroxy-3-methylglutaryl coenzyme A synthase mRNA,
DEFINITION complete cds.
ACCESSION L25798
VERSION L25798.1 GI:410027
KEYWORDS 3-hydroxy-3-methylglutaryl coenzyme A synthase.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE 1 (bases 1 to 1650)
AUTHORS Rokosz,L.L., Boulton,D.A., Butkiewicz,E.A., Sanyal,G., Cueto,M.A.,
Lachance,P.A. and Hermes,J.D.
TITLE Human cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase:
expression, purification, and characterization of recombinant
wild-type and Cys129 mutant enzymes
JOURNAL Arch. Biochem. Biophys. 312 (1), 1-13 (1994)
MEDLINE 94304197
PUBMED 7913309
COMMENT Original source text: Homo sapiens fetal adrenal cDNA to mRNA.
FEATURES
source location/Qualifiers
1..1650
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="fetal adrenal"
22..1584
/codon_start=1
/product="3-hydroxy-3-methylglutaryl coenzyme A synthase"
/protein_id="AAA6241.1"
/db_xref="GI:410028"
/translation="MPGSLPNTAAEACMPKDVGIIVALEIYPPSOYVDQAELEKXGVDA
GKTTIGLGQAKMGFCTDREDINSLCMTVVQNLMERNNLSYDCTGRLEVGTEETIIDSKS
SVKTNLMOLFEEBSGNTDIEGIDTNA CYGTA AFNAVNWISSSWDKRYALVVAADI
AVYATGNARPTGGVAVALLIGPNAPLIFERGLRGTMHQAYDFYKPDMLSEYPIVDG
KLSTQCYLSALDRCYSVYCKTHAQWKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLA
RMLNDFLNDQNRDKNSISGIEAFGDVKLEDTYFDRDVEKAFMKASSELEFSQKTKAS
LIVSNONGNMVTSVYSGIASVLAQYSPQQLAGKRIGVFSYSGLAATLYLSLVTQDA
TPGSALDKITASLCDLKSRLDBRTGVAPDVFAENMKLRREDTHHLVNYIPQGSIDSLFE
GTWYLVVRVDEKRRRTYARRPTBNDDTLDEGVGLVHSNIAIEHIPSAPKAVPRLPATTA
EPAAAVISNGEH"

ORIGIN
Alignment Scores:
Pred. No.: 1.52e-202 Length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 9 Gaps: 1

US-10-622-516-2 (1-478) x HUM3H3M (1-1650)
QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 22 ATGCCCTGATCACTTCCTTGAATGACAGAGCTTGCTGGCCAAAAGATGTGGGATTGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 82 GCCCTTGAGATCTATTTCCTTCTCAATATGTGATCAAGCAGAGAGTTGGAATAATATGAT 141

QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 142 GGTGTAGATGCTGGGAAAGTATACCAATGGCTTGCGCAGGCCAAGATGGGCTTCTGCACA 201
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATTAACTCTCTTTCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 262 AACCTTTCCTATGATTGCATTGGCGGCTGGAGITGGAAACAGACATCATCGACAAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAGTCTGTGAAGACTAATTGTGATGCAGCTGTTGAAGAGCTCGGGAATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 382 GAAGGATCGACACACTAATGCAATGCTATGAGGCACAGCTGCTGCTTCAATGCTGTT 441
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 442 AACTGAGTTGAGTCCAGCTCTTGGATGAGACGCTATGCCCTGTAGTTGACAGAGATATT 501
QY 149 ----- 149
Db 502 GCTGTATATGCCACAGGAATGCTAGACTACAGGTGAGATTGAGACAGTACTCTGCTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATGGGCCAAATGCTCTTAAATTTTGAACGAGGGCTTCGTGGACACATATGCCAACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 622 GCCTATGATTTTTCACAGCCTGATATGCTAATCTGAATATCTTATAGTAGAGGAAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 682 TCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTATCTGTCTACTGCAGAAAAGATC 741
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCCACTGGCAGAAAGAGGAAATGATAAAGATTTTACCTTGAATGATTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 802 ATGATCTTTCACTCACCACTAATTGTAACTGGTTCAGAAATCTTAGCTCGGATGTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCCTTAATGACCAAGATAGAGATAAATAAGATCTATAGTGGCTCGGAAGCC 921
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 922 TTTGGGATGTTAATTAGAAAGACACCTACTTTGATAGAGATGTGAGAGGCAATTATG 981
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 982 AAGGCTAGCTCTGAACCTCTTCAGTCAAGAAAACAAAGGCACTTTACTGTGATCAATCAA 1041
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1042 AATGAAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTCTAGCAGCAGTAC 1101
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1102 TCACCTCAGCAATTAGCAGGGAAGAAGATGGAGTGTCTTATAGTTCGTGGTTGGCT 1161
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1162 GCCACTCTGTACTCTCTTAAAGTCAACAMAGATGCTACACCGGGGTCTGCTCTGATATAA 1221

Qy	359	11eThrAlaSerLeuCysAspLeuLySerArgLeuAspSerArgThrGlyValAlaPro	378
Db	1222	ATTAACAGCAAGTTATGATGATTAATCAAGGCTTGATTCAGAACTGGTGGCACCA	1281
Qy	379	AspValPheAlaGluAsnMetLyLeuArgGluAspThrHisHisLeuValAsnTyrIle	398
Db	1282	GATGTCTTCGTGAATAACATGAAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT	1341
Qy	399	ProGInGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAspGlu	418
Db	1342	CCCCAGGGTTCAATAGATTCACTCTTGAAGAACGTGTACTAGTAGGGTGATGAA	1401
Qy	419	LyseHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGlyGly	438
Db	1402	AAGCACAGAAGAACTTACGCTCGGCTCCCACTCCAATGATGACACTTTGGATGAAGGA	1461
Qy	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysVal	458
Db	1462	GTAGGACTTGTCATTCAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA	1521
Qy	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
Db	1522	CCAAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAGCTGTGATTAGTAATGGGAACAT	1581
RESULT 4			
LOCUS	BC000297	2068 bp	mRNA linear PRI 30-SEP-2003
DEFINITION	Homo sapiens 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble), mRNA (cDNA clone IMAGE:2819708), partial cds.		
ACCESSION	BC000297		
VERSION	BC000297.2	GI:33991030	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2068)		
	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hejeh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Muliahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Vialalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butcherfield,Y.S., Krzywicki,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE	22388257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 2068)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	On Aug 20, 2003 this sequence version replaced gi:12653064.		
	Contact: MGC help desk		
	Email: cgapbs-r@mail.nih.gov		
	Tissue Procurement: DCTD/DTP		
	cDNA Library Preparation: Rubin laboratory		

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Lalic,P., Legaapi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantiriop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAL Plate: 1 Row: a Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504428.

FEATURES		location/Qualifiers	
source		1..2068	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:2819708"	
		/tissue_type="lung, small cell carcinoma"	
		/clone_lib="NIH MGC 7"	
		/lab_host="DH10B-R"	
		/note="Vector: pOTB7"	
gene		<1..2068	
		/gene="HMGCS1"	
		/note="synonym: HMGCS"	
		/db_xref="locusID:3157"	
		/db_xref="MIM:142940"	
CDS		<1..1668	
		/gene="HMGCS1"	
		/codon_start=1	
		/product="HMGCS1 protein"	
		/protein_id="AAH00297.2"	
		/db_xref="GI:33991031"	
		/db_xref="locusID:3157"	
		/db_xref="MIM:142940"	
		/translation="HSLSSAARSRLCPKEETVTDLETAVLYPSHSSFTMPGSLPLNAEACWPKDVGIVALEIYFPPSQYVDQAELEKYDGVADGKTYTIGLQAKMGFCTDREDINSLCMTVONLMERNNLSYDCIGLEVGTEETIIDKSKSVKTNLMQLFEESGNTDIEGIDPTNACYGCTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGCVGAVALLIGFNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDKLSIQCYLSALDRCSYVCKKHIAQWQKEGNDKDFITNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNISYSGLEAFGDVKLEDYFDRDVEKAFMKASSELFSQKTKASLLVSNONGNMYTSSVYGSLSAVLAQYSPQOLAGKRIGVFSYSGSLAATLYSLKVTQDAPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLRDFTHLNVNIPQGSIDSLFEGTWLYLVVDEKHRTRYARRPTPNDDTLDEGVGLVHSNIAATEHIPSPAKKVPRLPATPAEPEAAVISNGEH"	
misc_feature		142..1512	
		/gene="HMGCS1"	
		/note="HMG CoA synt; Region: Hydroxymethylglutaryl-coenzyme A synthase"	
		/db_xref="CDD:pfam01154"	
ORIGIN			
Alignment Scores:			
Pred. No.:		2.07e-202	
Score:		2480.00	
Percent Similarity:		91.92%	
Best Local Similarity:		91.92%	
Query Match:		98.77%	
DB:		9	
		Gaps: 1	
US-10-622-516-2 (1-478) x BC000297 (1-2068)			
Qy	1	MecProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLyAspValGlyIleVal	20
Db	106	ATGCCTGATCACTCTTTGAATGACAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT	165

QY	21	AlAlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGlnLysTyrAsp	40
Db	166	GGCCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCAGAGTTGAAAAATATGAT	225
QY	41	GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
Db	226	GGTGTAGATGCTGGAAGATATACCATGGCTTGGGCCAGGCCAAGATGGCTTCTGCACA	285
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValAlaGlnAsnLeuMetGluArgAsn	80
Db	286	GATAGAGAAGATATTAACCTCTTGTGCATGACTGTGTTCAGAACTCTTATGAGAGAAT	345
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
Db	346	AACTTTCTCTATGATTGCATTTGGCGGCTGGAAGTTGGAACAGAGCAATCATCGACAA	405
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGlnSerGlyAsnThrAspIle	120
Db	406	TCAAAGCTCTGTGAAGACTAATTGATGACAGCTGTTTGAAGAGCTCGGAATACAGATATA	465
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
Db	466	GAAAGGAATCGACACAACTAATGCATGCTATGAGAGCACAGCTGTCTTCAATGCTGTT	525
QY	141	AsnTrpIleGluSerSerSerTyrAsp-----	149
Db	526	AACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGTAGTTGCAGAGATATT	585
QY	149	-----	149
Db	586	GCTGTATATGCACAGAAATGCTAGACCTACAGGTGAGTTGAGCAGTAGCTTCGCTA	645
QY	150	-----GlyLeuArgGlyThrHisMetGlnHis	158
Db	646	ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGCTTCGTGGGACACATATGCAACAT	705
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu	178
Db	706	GCCTATGATTTTTCACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGAAAACTC	765
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
Db	766	TCCATACAGTGTCACTCACTAGTGCATTAGACCCGCTGCTACTCTGTCTACTGCAAAAAGATC	825
QY	199	HisAlaGlnTyrGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
Db	826	CATGCCCAGTGCAGAAAGAGGAAATGATAAAGATTTTACCCTTGAATGATTTTGGCTTC	885
QY	219	MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu	238
Db	886	ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTTACGCTCGGATGTTGCTG	945
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla	258
Db	946	AATGACTTCTTAATGACCAGATAAGATATAAAATAGATCTATAAGTGCGCTGGAAGCC	1005
QY	259	PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet	278
Db	1006	TTTGGGAGTTTAAATTAGAAGACACCTACTTTGATTAGAGATGAGGAAGGCATTATAG	1065
QY	279	LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln	298
Db	1066	AAGGCTAGCTCTGAACCTCTTCAGTCAAGTCAAAAAAAGGCATCTTACTTGATCAAAATCAA	1125
QY	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
Db	1126	AATGGAATAATGTACACACATCTTCAGTATATGTTCCCTTGCACTGTCTTACAGACAGTAC	1185
QY	319	SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
Db	1186	TCACTCAGCAATTAGCAGGGAAGAAGATGAGAGTGTTCCTTATGCTTGGTTTGGCT	1245

Qy		339	AlaThrLeuTyrSerLeuIysValThrglnAspAlaThrProGlySerAlaLeuAspLys	358
Db		1246	GCCACTCTGTACTCTCTTAAGTCACACAAGATGCTACACCGGGGTCTGCTTTGATATAA	1305
Qy		359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
Db		1306	ATAACAGCAAGTTATTATGTGATCTTAATCAAGCCTTGATTCAAGAAGCTGGTGCACCA	1365
Qy		379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleuValAsnTyrrile	398
Db		1366	GATGCTTCGCTGAANAACATGAAGCTCAGAGAGAACACCACCATCATTTGGTCAACTATATT	1425
Qy		399	ProGInglySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu	418
Db		1426	CCCCAGSGTTCAATAGATTCACTCTTTGAAGGAACGTGTACTTAGTTAGGGTGAATGAA	1485
Qy		419	LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly	438
Db		1486	AAGCACAGAAGACTTACGCTCGGCGCTCCCACTCCAATGATGACACTTTGGATGAAGGA	1545
Qy		439	ValGlyLeuValHisSerAsnIleAlaThrGlnHisIleProSerProAlaLysLysVal	458
Db		1546	GTAGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAAGCCCTGCCAAGAAAGTA	1605
Qy		459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGlnHis	478
Db		1606	CCAAGACTCCCTGCCACAGCAGACAGAACTGAAGCAGCTGTCTATTAGTAATGGGAACAT	1665
RESULT 5				
HSCoAS				
DEFINITION	H.sapiens mRNA for HMG-CoA-synthase.	1685 bp	mRNA	linear PRI 12-SEP-1993
ACCESSION	X66435 S48133			
VERSION	X66435.1 GI:30008			
KEYWORDS	Hydroxymethylglutaryl CoA Synthase.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Russ,A.P., Ruzicka,V., Maerz,W., Appelhaus,H. and Gross,W.			
TITLE	1 (bases 1 to 1685)			
JOURNAL	Amplification and direct sequencing of a cDNA encoding human cytosolic 3-hydroxy-3-methylglutaryl-coenzyme A synthase Biochim. Biophys. Acta 1132 (3), 329-331 (1992)			
MEDLINE	93041939			
PUBMED	1358203			
REFERENCE	2 (bases 1 to 1679)			
AUTHORS	Russ,A.			
TITLE	Direct Submission			
JOURNAL	Submitted (26-MAY-1992) A. Russ, Labor fur angewandte Biochemie, Theodor Stern-Kai 7, W-6000 Frankfurt am Main 70, FRG			
REMARK	revised by [2]			
REFERENCE	3 (bases 1 to 1685)			
AUTHORS	Russ,A.			
TITLE	Direct Submission			
JOURNAL	Submitted (10-AUG-1992) Andreas Russ, Zentrum der biologischen Chemie, J.W.-Goethe-Universitaet Frankfurt, Theodor-Stern-Kai 7, Frankfurt, 6000, Germany			
FEATURES	Location/Qualifiers			
source	1..1685			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="HMGCoASynthase"			
	/cell_type="Fibroblast"			
	123..1685			
	/codon_start=1			
	/product="Hydroxymethylglutaryl CoA Synthase"			
	/protein_id="CAA47061.1"			
	/db_xref="GI:30009"			
	/db_xref="GOA:Q01581"			
	/db_xref="SWISS-PROT:Q01581"			
	/translation="MPSGLPLNAEACWPKDVGIVALEIFYPSQYVDQALEKYGVDVA			
CDS				

KYITIGLQAKMGFCTDREDINSLCMTVQNLMERNNLSYDCIGRLEVEETETIIDSKS
SVKTNLMQLFEESGNTDIEGIDDTNACYGGTAAFVNAVNWIESSWDGRYALVAGDI
AVYATGNARPTGGVAVALLIGPNAPLI FERGLRTHQHDYDFYKPDMLSEYPTVDG
KLSIQCYSALDRDCYSVYCKIHAQMÖKEANDNDFTLNDFGFMI FHSPLYCKLVOKSLA
RMLINDFLNDONRDKNISYSGLKAFGDVKLEDYTPDRDVEKA FMKASSELFSOKTKAS
LLVSNÖNGMNTSSVYGSILASVLAQYS PÖHLAKRIGVFSYSGSLAATLYSLKVTQDA
TPGSALDKITASLCDLKRSLDRSTGVAQDVFAENMKLRBDTHLVNI PÖGSIDSLFE
GTWYLVVRVDEKHRRTYARRPTPNDDTLDEGVGLVHSNIATEHIIPSPAKKVPRLPATAA
EPEAAVISNGVM"

ORIGIN

Alignment Scores:

Pred. No.: 5.04e-199 Length: 1685
Score: 2439.00 Matches: 471
Percent Similarity: 91.12% Conservative: 1
Best Local Similarity: 90.93% Mismatches: 4
Query Match: 97.13% Indels: 42
DB: 9 Gaps: 1

US-10-622-516-2 (1-478) x HSCoAS (1-1685)

QY	1	MePProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLyAspValGlyIleVal	20
DB	123	ATGCTTGATCACTTCCTTGAATGAGAAGCTTGCTGGCCAAAGATGTGGAAATTGTT	182
QY	21	AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp	40
DB	183	GCCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGAGTTGGAATAATATGAT	242
QY	41	GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
DB	243	GGTGTAGATGCTGGAAGATATACCAATGGCTTGGCCAGGCCAAGATGGCTTGCACA	302
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn	80
DB	303	GATAGAGAAGATATTAACTCTTTCGATGACTGTGTTCAAGATCTTATGAGAAGAAAT	362
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
DB	363	AACCTTTCCTATGATTGCATTGGCGGCTGGAAGTTGGAACAGACAATCATCGACAAA	422
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
DB	423	TCAAAGTCTGTGAAGACTAATTGTATGCAGCTGTTGAAGAGTCTGGGAATACAGATATA	482
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
DB	483	GAAAGGATCGACACAACAATAATGCTATGGAGGCACAGCTGCTGTTCAATGCTGTT	542
QY	141	AsnTrpIleGluSerSerSerTrpAsp-----	149
DB	543	AACGTGATTGAGTCCAGCTCTTGGAATGGAAGGTATGCCCTGTAGTTGCAGAGATATT	602
QY	149	-----	149
DB	603	GCTGTATATGCCACAGGAATGCTAGACCTACAGGTGAGTTGGACAGTAGCTCTGCTA	662
QY	150	-----GlyLeuArgGlyThrHisMetGlnHis	158
DB	663	ATTGGGCCAAATGCTCCTTAATTTTGGAACGAGGGCTTCGTGGACACATATGCAACAT	722
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu	178
DB	723	GCCATATGATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGATGAAAACTC	782
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
DB	783	TCCATACAGTGTACTACCTCAGTCATTAGACCGCTGCTATTCTGTCTACTGCAAAAAGATC	842
QY	199	HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
DB	843	CATGCCCACTGGCAGAAAGGCAAAATGATTAACGATTTTTACCTTGAAATGATTTGGCTTC	902

QY	219	MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu	238
DB	903	ATGATCTTTCACTCACCACCATATTGTAAACTGGTTCAGAAATCTTAGCTCGGATGTGCTG	962
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla	258
DB	963	AATGACTTCTTAATGACCAGATAAGATAAAATAATAGTATCTATAGTGCGCTGAAGGCC	1022
QY	259	PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet	278
DB	1023	TTTGGGATGTTAAGTTAGAGACACCTACTTTGATAGAGATGTGAGAAGCATTTATG	1082
QY	279	LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln	298
DB	1083	AAGGCTAGCTCTGAACCTCTTCACTCAGAAAACAAAGGCATCTTTACTTGATCAATCAA	1142
QY	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
DB	1143	AATGAAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTCTAGCACAGTAC	1202
QY	319	SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
DB	1203	TCACCTCAGCATTTAGCAGGGAAGAATTGGAGTGTTTCTTATAGTTCTGTTGGCT	1262
QY	339	AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys	358
DB	1263	GCCACTCTGTACTCTCTTAAAGTACACAAGATGCTACACCGGGGCTGCTCTGTATAA	1322
QY	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
DB	1323	ATAACGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTGTGTGCACAA	1382
QY	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle	398
DB	1383	GATGCTTCGCTGAAAAACATGAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT	1442
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu	418
DB	1443	CCCCAGGTTCAATAGATTCACTCTTGAAGGAACGTGTACTTAGTTAGGTGATGAA	1502
QY	419	LysHisArgArgThrTyrAlaArgArgProThrProAsnAspThrLeuAspGluGly	438
DB	1503	AAGCAGAGAAGACTTAGCGCTGGGCTCCCACTCCAATGATGACACTTGGATGAAGA	1562
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal	458
DB	1563	GTAGGACTTGTGCATTCAAAACATAGCAACTGAGCATATTCCAAAGCCCTGCCAAGAAAGTA	1622
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGly	476
DB	1623	CCAAGACTCCCTGTACAGCAGCAAACTGAAGCAGCAGTTATTAATGATGGG	1676
RESULT 6			
AK095492			
LOCUS			
DEFINITION			
Homo sapiens cDNA FLJ38173 fis, clone FCBBF1000053, highly similar			
to HYDROXYMETHYLGLUTARYL-COA SYNTHASE, CYTOPLASMIC (EC 4.1.3.5).			
AK095492			
ACCESSION			
AK095492.1 GI:21754757			
KEYWORDS			
oligo capping; fis (full insert sequence).			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE			
AUTHORS			
1 Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,			
Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,			
Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J.,			
Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K.,			
Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H.,			
Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A.,			
Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuno, Y., Nagai, K.			
and Isogai, T.			

TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3195)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB) ; cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

FEATURES
source location/Qualifiers
1..3195

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="FCBF1000053"
/tissue_type="brain"
/clone_lib="FCBF1"
/dev_stage="fetal"
/note="Cloning vector: pME18SFL3"
68..1597
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC04559.1"
/db_xref="GI:21754758"
/translation="MEGSLPLNAEACWPKDVGIVALEIYPPSQYVDQAELEKYDGVDA
GKYTIGLQAKMGFCTDREDINSLCMTVVQNLMERNNISYDCIGLEVGTEIIDS
SVKTNLMQLFEESGNDIEGIDTNAVNWIESSWDGRYALVAGDIAVYATGNARPT
GGVGVALLIGPNAPLIEFGRGTHQHA YDFYKPMLESPYVDGLSIOCYLSAL
DRCVSYCKKIHAQMOKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQ
NRDKNSIYSGLEAFGVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGMY
TSSVYGSILASVLAQYSPQQLAGKRIGVSYSGSLAATLYSLKVTQDATPGSALDKITA
SLCDIKSRLDSRTGVAPDVFAENMKUREDTHLVNYTPQGSIDSLFEGTWYLVRVDEK
HRRFYARRPTFNDPTDEGVGLVHSNIAIEHIIPSPAKKVPRLPATAEPEAAVISNEE
H"

CDS
ORIGIN
Alignment Scores:
Pred. No.: 1.3e-195 Length: 3195
Score: 2403.50 Matches: 467
Percent Similarity: 89.81% Conservative: 0
Best Local Similarity: 89.81% Mismatches: 0
Query Match: 95.72% Indels: 53
DB: 9 Gaps: 2

US-10-622-516-2 (1-478) x AK095492 (1-3195)
QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 68 ATGCCTGATCATCTTCCTTGAAATGCAGAAAGCTTGCGCCAAAGATGTGGGAATTGTT 127
QY 21 AlaleuGluIleTyrPheProSerGlnTyrValaspGlnAlaGluLeuGluLysTyrAsp 40
Db 128 GCCCTTGAGATCTATTTCTCTCAATATGTGATCAAGCAGAGTTGGAAAAATATGAT 187
QY 41 GlyValaspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 188 GGTGTAGATGCTGGAAGATATACCATTTGGCTGGCCAGGCCAAGATGGGCTTCTGCACA 247

QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 248 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGTTCAAGAACTTATGAGAGAAGAAAT 307
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 308 AACCTTCTATGATTGATTTGGGCGGCTGGAAGTTGGAAACAGACAAATCATCGACAAA 367

QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 308 AACCTTCTATGATTGATTTGGGCGGCTGGAAGTTGGAAACAGACAAATCATCGACAAA 367

QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 368 TCAAAGCTGTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGCTGTGGAAATACAGATATA 427
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 428 GAAGGATCGACACAACACT-----AATGCTGTT 454
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 455 AACTGGATTGAGTCCAGCTCTTGGGATGACGGTATGCCCTGTAGTTGCAGAGATATT 514
QY 149 ----- 149
Db 515 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGAGCAGTAGCTTGCTTA 574
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 575 ATGGGCCCAAATGCTCCTTTAATTTTGAACGAGGGCTTGTGGACACATATGCAACAT 634
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 635 GCTATGATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGAAACTC 694
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 695 TCCATACAGTGTCACTCAGTGCATTAGACCCGCTGCTATTCTGTCTACTGCCAAAAGATC 754
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 755 CATGCCACAGTGGAGAAAGAGGAATGATTAAGATTTTACCTTGAATGATTTGGCTTC 814
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeu 238
Db 815 ATGATCTTTCACCTCACCATATGTGAAACTGGTTCAGAAATCTAGCTCGAGTGTGCTG 874
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 875 AATGACTTCTTAATGACCAGCAATAGATAAATAATAGTATCTAATGCGCTGGAAGCC 934
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 935 TTGGGGATGTTAAATTAGAGACACCTACTTTGATAGAGATGGAGAGGCAATTATG 994
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 995 AAGGCTAGCTCTGAACCTCTTCACTCAGAAACAAAGGCATCTTTACTTGATCAAAATCAA 1054
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1055 AATGGAATATGTACACATCTTCAGTATATGGTTCCCTTGCACTGTTCTAGCACAGTAC 1114
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1115 TCACCTCAGCAATTAGCAGGGAAGAATGGAGTGTCTTATGTTCTGTGGCT 1174
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1175 GCCACTGTGTACTCTCTTAAAGTCAACAGATGCTACACGGGGTCTGCTCTGATAAA 1234
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1235 ATAAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTGTGTGGCACA 1294
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db 1295 GATGTCTTGGTGAAGAAACATGAAGCTCAGAGAGACACCATCATTTGGTCAACTATATT 1354
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAspGlu 418
Db 1355 CCCCAGGGTTCAATAGATTCACCTTTTGAAGGAACGTGTACTTAGTGAAGGTGGATGAA 1414
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438

Db 1415 AAGCAGAGAAGACTTACGCTCGGCGTCCCACTCCCAATGATGACACTTTGGATGAAGGA 1474

QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysVal 458

Db 1475 GTAGGACTTGTCATTCAACACTAGCAACTGAGCATATTCCAAAGCCCTGCCAAGAAAGTA 1534

QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478

Db 1535 CCAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAGCTGTCAATTGATGAGGAACAT 1594

RESULT 7

AX700129 3275 bp DNA linear PAT 03-APR-2003

LOCUS AX700129

DEFINITION Sequence 15 from Patent EP1284298.

ACCESSION AX700129

VERSION AX700129.1 GI:29536069

KEYWORDS

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1

AUTHORS Brooksbank, R.A., Dixon, A.K., Lee, K. and Pinnock, R.D.

TITLE Identification and use of molecules implicated in pain

JOURNAL Patent: EP 1284298-A 15 19-FEB-2003;

WARNER-LAMBERT COMPANY (US)

FEATURES

source 1..3275

/organism="Rattus norvegicus"

/mol_type="unassigned DNA"

/db_xref="taxon:10116"

/note="Cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase"

ORIGIN

Alignment Scores:

Pred. No.: 4.77e-192 length: 3275

Score: 2362.00 Matches: 455

Percent Similarity: 89.62% Conservative: 11

Best Local Similarity: 87.50% Mismatches: 12

Query Match: 94.07% Indels: 42

DB: 6 Gaps: 1

US-10-622-516-2 (1-478) x AX700129 (1-3275)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20

Db 92 ATGCTGGGTCACTTCCTTGATGACAGAGCTTGCTGGCCAAAGATGTGGGAATCGTT 151

QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40

Db 152 GCCCTTGAATCTACTTTCCTCTCAAGTATGTTGATCAAGCTGAGTTGAAAAATACGAT 211

QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60

Db 212 GGTGTGATGCTGGAAAGTATACATTGGCCCTGGCCAGGCCAGATGGGCTTCTGCACG 271

QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80

Db 272 GATCGCGAAGACATCAACTCTCTGCTGCTGACTGTGTTCAGAAACTGATGAGAAAT 331

QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100

Db 332 AGCCTTCTCTACTGATGCAATGGGGCGGTGGAAGTCGGAGCAGACAAATCATGCACAA 391

QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120

Db 392 TCAAAATCCGTGAAGTCTAATTGATGACAGCTGTTGAGGAGCTCGGAATACAGATATA 451

QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140

Db 452 GAAGGAATAGATACAACTAATGCATGCTATGGGGGCACAGCCGCACTTCAATGCTGTG 511

QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149

Db 512 AACTGATCGAATCCAGCTCTTGGATGACGATACGCTTGGTAGTTGCAGAGACATC 571

QY 149 ----- 149

Db 572 GCTATATATGCCCTCAGGAACGCCAGCCTACAGGTGAGTTGAGCTGTGCTGTGCTA 631

QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158

Db 632 ATGGGCCCAAATGCTCCTGTAAATTTTGAACCGAGGGCTTCTGGGACACACATGCAGCAT 691

QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178

Db 692 GCCTACGACTTTACAAAGCCTGACATGCTCTCTGAATACCTGTGTAGATGGAACATC 751

QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198

Db 752 TCCATACAGTGTACTACCTCAGCGCATTTGACCGCTGTATCTGTCTACCGCAAAAAGATC 811

QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218

Db 812 CGGGCCCAAGTGGCAGAAAGAGGAAAGATAAAGATTTCACCTGAATGATTTGGCTTC 871

QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238

Db 872 ATGATCTTCACTCGCCCACTACTGTAAACTGGTGCAAGAAATCTCTAGCTAGATGTTCCCTG 931

QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258

Db 932 AATGACTTCTTAAAGATCAAAACAGACAAAAACAGATTTTACAGTGGGCTGGAAGCC 991

QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278

Db 992 TTTGGGAGTGTGAATTAGAAGATACTTACTTCGACAGAGATGTGAAAAAGCATTTATG 1051

QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298

Db 1052 AAGGCTAGTGTGAGCTATTCACCAAGAAACAAAGGCATCTTGTGTTATCGAATCAA 1111

QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318

Db 1112 AATGAAACATGTACACATCTCTGTATACGGTTCCTTGCTTCTTGCCACAGTAC 1171

QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338

Db 1172 TCACCTCAACAGTTGGCCGGGAAGAGATTGAGTGTCTTCAAGGTTCTGGCTTGCT 1231

QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358

Db 1232 GCCACACTCTACTCCCTTAAAGTACACACAAGATGCCACACAGATCTGCTTGACAAA 1291

QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378

Db 1292 ATAACAGCAAGTTATGTGACCTTAAGTCAAGGCTTGACTCAAGAACGTGTGGCACCA 1351

QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle 398

Db 1352 GACGCTTTGCTGAAAACATGAACTCAGAGAGCACACATCACTTGAACCACTATATT 1411

QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAspGlu 418

Db 1412 CCCCAAGTTTCAATAGATTCACTCTTCGAAGGAACAATGTAATCTAGTCAAGTGGATGA 1471

QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438

Db 1472 AAGCAGAGAAGAACTTACGCCGGCGTCCCTCCACAAATGACCAAGTTTGATGAAGGA 1531

QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458

Db 1532 GTGGGACTTGTGATTTCAACACACAGCTACAGAGCATATTCCAAGCCCGGCTAAGAAAGTG 1591

QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db 1592 CCAAGACTTCCTCGCAGCTCGGGCGAAGCTGAGTCGGCTGTCAATCAAGTAACGGGAGAC 1651
RESULT 8
AX827303
LOCUS AX827303 3275 bp DNA linear PAT 12-DEC-2003
DEFINITION Sequence 37 from Patent EP1344834.
ACCESSION AX827303
VERSION AX827303.1 GI:39837392
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
AUTHORS Boess, F., Suter-Dick, L. and Wolf, D.
TITLE Methods for the toxicity prediction of a compound
JOURNAL Patent: EP 1344834-A 37 17-SEP-2003;
F. HOFEMANN-LA ROCHE AG (CH)
FEATURES
SOURCE location/Qualifiers
1.3275
/organism="Rattus norvegicus"
/mol_type="unassigned DNA"
/db_xref="taxon:10116"
ORIGIN
Alignment Scores:
Pred. No.: 4.77e-192 Length: 3275
Score: 2362.00 Matches: 455
Percent Similarity: 89.62% Conservative: 11
Best Local Similarity: 87.50% Mismatches: 12
Query Match: 94.07% Indels: 42
DB: 6 Gaps: 1
US-10-622-516-2 (1-478) x AX827303 (1-3275)

Db 572 GCTATATATGCTTCAGGAAACGCCAGCCTACAGGTGGAGTTGAGCTGTGCTTCCTA 631
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 632 ATTGGGCCCAATGCTCTCTGTAATTTTGGACCGAGGGCTTCGTGGGACACACATGACGAT 691
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 692 GCCTACGACTTTTACAAGCCTGACATGCTCTGTAATACCCTGTGCTAGATGAGAAAAC 751
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIle 198
Db 752 TCCATACAGTGTACCTCAGCGCATTTGACCGCTGCTATTCGTCTACCGCAAAAGATC 811
QY 199 HisAlaGlnTyrGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 812 CCGGCCAGTGGCAGAAAGAGGAAAGATTAAGATTTTACCCTGAATGATTTGGCTTC 871
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeu 238
Db 872 ATGATCTTTCACTCGCCATACGTAAACTGTGCAGAAATCTCTAGCTAGATGTTCTCTG 931
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 932 AATGACTTTCTTAACGATCAAAACAGAGACAAACAGATTTTACGTGGCTGGAAGCC 991
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 992 TTTGGGATGTGAATTTAGAGATACTTACTTCGACAGAGATGTGAAAAGCATTTATG 1051
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 1052 AAGGCTAGTGTGAGCTATTCACAGAAACAAAGGCATCTTTGCTGTATCGAATCAA 1111
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1112 AATGAAACATGTACACATCTCTGTATACGTTCCCTGCTTCTGTCTGACACAGTAC 1171
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1172 TCACCTCAACAGTTGGCCGGGAAAGAGATGAGTGTCTCTTACGTTCTGGCTTGCT 1231
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1232 GCCACACTCTACTCCCTTAAAGTCACACAGATGCCACACAGATCTGCTTGACAAA 1291
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1292 ATACAGCAAGTTATGTGACCTTAAGTCAAGGCTTGACTCAAGAACGTGTGGCACCA 1351
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisLysLeuValAsnTyrIle 398
Db 1352 GACGCTTTGCTGAAAACATGAAGCTCAGAGAGACACACATCACTTAAGCCAATATATT 1411
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAspGlu 418
Db 1412 CCCAGTGTCAATAGATTCACTCTTGAAAGGAACATGTATCTAGTCAAGTGGATGAA 1471
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1472 AAGCAGAGAAAGACTTACGCCCGCGCTCCCAACAATGACACACAGTTGGATGAAGA 1531
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db 1532 GTGGGACTTGTGATTCAAACACAGCTACAGACATTTCCAAAGCCGGCTAAGAAAGTG 1591
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db 1592 CCAAGACTTCCTCGCAGCTCGGGCGAAGCTGAGTCGGCTGTCAATCAAGTAACGGGAGAC 1651
RESULT 9
RNCHMCA 3275 bp mRNA linear ROD 18-JUN-2003
LOCUS RNCHMCA

DEFINITION Rattus norvegicus mRNA for cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase (EC 4.1.3.5) .
ACCESSION X52625
VERSION X52625.1 GI:55946
KEYWORDS cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1
AUTHORS Ayte,J., Gil-Gomez,G. and Hegardt,F.G.
TITLE Nucleotide sequence of a rat liver cDNA encoding the cytosolic 3-hydroxy-3-methylglutaryl coenzyme A synthase
JOURNAL Nucleic Acids Res. 18 (12), 3642 (1990)
MEDLINE 90301491
PUBMED 1972979
REMARK (b 1-38,40,41-62,64-3275)
REFERENCE 2
AUTHORS Hegardt,F.G.
TITLE Direct Submision
JOURNAL Submitted (17-APR-1990) Hegardt F.G., University of Barcelona, Unit of Biochemistry, School of Pharmacy, Placa Pius XII, s/n. 08028 barcelona, Spain
REMARK (b 1-38,40,41-62,64-3275)
REFERENCE 3 (bases 1 to 3275)
AUTHORS Hegardt,F.G.
TITLE Direct Submision
JOURNAL Submitted (30-JUL-1990)
FEATURES
source
1..3275
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague Dawley"
/db_xref="taxon:10116"
/clone="lambda-CCS1"
/tissue_type="liver"
31..70
/note="revised by [3]"
/citation=[3]
92..1654
/EC_number="4.1.3.5"
/codon_start=1
/product="cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase"
/protein_id="CAA36852.1"
/db_xref="GI:55947"
/db_xref="GOA:P17425"
/db_xref="SWISS-PROT:P17425"
/translation="MPGSLPLNABACWPKDVGIVALEIYFPSQYVDAQAELEKYGVDVAGKYTIGLGQARMGFCITREDINSLCLTVQKLMERNLSYDCIGRLVEGTETIIDSKSVKSNLMQLFEESGNTDIEGIDTINACYGTAAFNVAVMWIESSMDGRVALVAGDI AIYASGNARPTGGVGAVALLIGPNAPVIFDRGLRGTHQAAYDFYKPDMLSEYPVDG KLSIQCYLSALDRCYSVYRKIRAQWQKEGKDKDFTLNDFGFMI FHSPYCKLVQKSLA RMFLNDPLNDQNRDKNSISGLEAFGDKLEDITYFDRDVEKAFMKASAELEFNQTKAS LLSVNQNGMNTSSVYGSILASVLAQYSPQQLAGKRIGVFSYSGSLAATLISLKYTQDA TPGSALDKITASLCDLKSRLDSRTCVA PDVFAENMKLRDTHILANYI POCSTDSLFE GTWYLVRVDEKHKRTYARBPSTNDHSLDEGVGLVHSNTATEHIPS PAKKVPRLPATSG EPESAVISNGEH"

ALIGNMENT SCORES:
Pred. No.: 4.77e-192 Length: 3275
Score: 2362.00 Matches: 455
Percent Similarity: 89.62% Conservative: 11
Best Local Similarity: 87.50% Mismatches: 12
Query Match: 94.07% Indels: 42
DB: 10 Gaps: 1

US-10-622-516-2 (1-478) x RNCMCA (1-3275)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProIysAspValGlyIleVal 20
|||||

Db 92 ATGCCTGGGTCACTTCCTTGAATGCAGAGGCTTGCTGGCCAAAGATGTGGAAATCGTT 151
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluIlySerTyrAsp 40
Db 152 GCCCTTGAATCTACTTTCCTTCTCAGTATGTGATCAAGCTGAGTTGAAAAATACGAT 211
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 212 GGTGTAGATGCTGGAAAGTATACCATTTGGCCTGGCCAGGCCAGAGTGGCTTCTGCACG 271
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 272 GATCGCGAAGACATCAACTCTCTGCTGCTGACTGTGGTTGAGAAACTGATGAGAGAAAT 331
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleLeuAspLys 100
Db 332 AGCCTTCTATGACTGCATTGGCGGCTGGAAGTCGGGACAGACAATCATCGACAA 391
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGlnGluSerGlyAsnThrAspIle 120
Db 392 TCAAAATCCGTGAAGTCTAATTGATGCAGCTGTTTGAGGAGTCTGGGAATACAGATATA 451
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 452 GAAGGAATAGATACAACTAATGATGCTATGGGGCAGACGCCAGCTTCAATGCTGTG 511
QY 141 AsnTrpIleGluSerSerSerTyrAsp----- 149
Db 512 AACTGATCGAATCCAGCTCTGGGATGAGCATACGCTTGGTAGTTGCAGAGACATC 571
QY 149 ----- 149
Db 572 GCTATATATGCCTCAGGAACGCCAGCCTACAGGTGAGTTGAGCTGTGCTGCTA 631
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 632 ATTGCGCCAAATGCTCCTGTAAATTTTGAACCGAGGCGCTTGTGGGACACACATGCAGCAT 691
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 692 GCCTACGACTTTTACAAGCCTGCATGCTCTCTGAATACCTGTGTGATGAGAAAACTC 751
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 752 TCCATACAGTGTACCTCAGCGCATTTGGAACCGCTGCTATTCGTCTACCGCAAAAAGATC 811
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 812 CGGGCCCACTGGCAGAAAGAGGAAAGCATTAAGATTTCACCTGAATGATTTGGCTTC 871
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeu 238
Db 872 ATGATCTTTCACCTGCCATACCTGTAACTGTGCAGAAATCTCTAGACTAGGATGTTCCCTG 931
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 932 AATGACTTCTTAACGATCAAAACAGAGACAAAACAGTATTACAGTGGGCTGGAAGCC 991
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 992 TTTGGGATGTGAATTAGAAGATACTTACTTCGACAGAGATGTGAAAAAGCATTTATG 1051
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 1052 AAGGCTAGTGTGAGCTATTCAACAGAAAACAAAGGCATTTTGCTGTATCGAATCAA 1111
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1112 AATGGAACATGTACACATCTCTGTATACGGTTCCCTTGCTTCTGTGCACAGTAC 1171
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1172 TCACCTCAACAGTTGGCCGGAGAGGATTGGAAGTGTCTTCAAGGTTCTGGCTTGCTGCT 1231

QY	339	AlaThrLeuTyrSerLeuLySerValThrGlnAspAlaThrProGlySerAlaLeuAspLys	358
Db	1232	GCCACACACTTACTCCCTTAAAGTCACACAGAGTCCACACAGATCTGCTTGACAAA	1291
QY	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
Db	1292	ATAACAGCAAGTTTATGTGACTTTAAGTCAAGGCTTGACTCAGAACGTTGTGGCACCA	1351
QY	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle	398
Db	1352	GACGCTTTGCTGAAAACATGAAAGCTCAGAGAGACACACATCACTTAGCCACTATATT	1411
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu	418
Db	1412	CCCCAGTGTCAATAGATTCACTCTTGAAGGAACATGATCTAGTCAGAGTGGATGAA	1471
QY	419	LySHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly	438
Db	1472	AAGCACAGAGAAGACTTACGCCGGCGCTCCCTCCACAATGACCAAGTTTGATGAAGGA	1531
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGlnHisIleProSerProAlaLysLysVal	458
Db	1532	GTGGGACTGTGCATTCAAAACACAGCTACAGACATATTCCAAGCCCCGGCTAAGAAAGTG	1591
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
Db	1592	CCAAAGACTTCTGCGACCTCGGGCGAACTGAGTGGCTGTCACTCACTAAACGGGAGACAC	1651
RESULT 10			
LOCUS	BC023851	2703 bp	mRNA linear ROD 06-OCT-2003
DEFINITION	Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA		
ACCESSION	BC023851		(CDNA clone MGC:36620 IMAGE:5347038), complete cds.
VERSION	BC023851.1	GI:23271493	
KEYWORDS	MGC.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 2703)		
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Uedin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kerteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Buterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalhus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE	22388257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 2703)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		

COMMENT

Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gailthersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hgrl.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Latic,P., Legaspi,R., Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantiripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAK Plate: 55 Row: h Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22122370.

FEATURES

source

1..2703
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:36620 IMAGE:5347038"
/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal carcinoma, 5 month old virgin mouse."
/clone_lib="NCI CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..2703

gene

/gene="Hmgcs1"
/note="synonym: MGC36662"
/db_xref="locusID:208715"
/db_xref="MGI:107592"
149...1711

CDS

/codon_start=1
/product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1"

/protein_id="AAH23851.1"
/db_xref="GI:23271494"
/db_xref="locusID:208715"
/translation="MPGSLPLNAEACPKDVGIVALEITYPPSQYVDQAELEKYDGVDA GKXTIGIQARMGFTCDIEDINTNCLTVQKLMERHLSYDCIGRLVEGETIIDKSK SVKSNLMQLFEESGNTDIEGIDNTNACYGSTAAVFNVMVESSSWDGRYALVAGDI AIVATGNARPTGSGVAVALLIGPNAPLIFDRGLRGTHQHAIDFYKPPMLSEYPVVDG KLSIQCYLSALDRCSVYRKIRAQWQKEGKDKDFTLNDFGFMI FHSPPCKLVQKSLA RMFLNDFLNDQNRDKNSIYSGLEAFGVDKLEDITYFDRIEKA FMKASSEIFNQTKAS LTVSNQNGNMTYSVYGSILASVLAQYSPQLAGKRVGSYSGLAATLYSLKVTODA TPGSALDKITASLCDKSLRDSRTCAPDVFAENMKREDTHLANIYIQCSIDSLFE GTWYLVVRVDEKHRRTYARRPFTNDHSLDEGMGLVHSHNTATEHIPSAPKVPRLPATSA ESESAVISNGEH"
185...1555
/note="HMG CoA synt; Region: Hydroxymethylglutaryl-coenzyme A synthase"
/db_xref="CDD:pfam01154"

ORIGIN

misc_feature

/note="HMG CoA synt; Region: Hydroxymethylglutaryl-coenzyme A synthase"
/db_xref="CDD:pfam01154"

Alignment Scores:

Pred. No.:	5,8e-191	Length:	2703
Score:	2348.00	Matches:	452
Percent Similarity:	89.42%	Conservative:	13
Best Local Similarity:	86.92%	Mismatches:	13
Query Match:	93.51%	Indels:	42
DB:	10	Gaps:	1

US-10-622-516-2 (1-478) x BC023851 (1-2703)

QY 1 MecProGlySerLeuProLeuAenAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 149 ATGCTGGGTCACTTCCTTGAATGCAGAGGCTTGCGCCAAAAGATGTGGGAATCGTT 208
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 209 GCCCTTGAATCTACTTTCCTTCTCAATATGTGATCAAGCTGAGTTGAAAAATACGAT 268
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 269 GGTGTAGATGCTGGAAGATATACCATCGCGCTGGCCAGGCCAGATGGGCTTCTGACG 328
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 329 GATCGTGAAGACATCAACTCTCTTGCCCTGACTGTGTTCAGAAACTGATGAGAGACAT 388
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 389 AGCCTTTCCTATGATGATGCATTTGGCGGCTAGAAGTTGAAACAGACATCATCGACAA 448
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 449 TCGAAATCAGTGAAGTCTAATTTGATGCAAGCTGTTGAGGAGTCTGGGAATACAGATATA 508
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 509 GAAGGATAGATACAACCAATGCAATGCTATGGGGCACAGCTGCACTTCAATGCCGTG 568
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 569 AACTGGGTCGAATCCAGCTCTTGGAATGACGATATGCTGTGTAAGTTGCAGAGACATT 628
QY 149 ----- 149
Db 629 GCTATATATGCCACAGAAATGCCAGACCTACAGGTGAGTTGAGCTGTGCCCTGCTA 688
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 689 ATTGGGCCAAACGCTCCTTAATTTTGAACCGAGGGCTCCGTGGGACACACATGCAGCAT 748
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 749 GCCTATGACTTTTACAAAGCCTGACATGCTCTCCGAGTACCCTGTGTCGACGGAAGCTC 808
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 809 TTCATATACAGTGTACTACCTCAGCGCCCTGGAACCGCTGCTATTCTGTACCGCAAAAAGATC 868
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 869 CGTGCCCAAGTGGCAGAAAGAGGAAAGATAAAGATTTTACCCTGAATGATTTTGGCTTC 928
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 929 ATGATCTTTCACTCACCATATTTGTAAACTGCTGCAGAAATCTTAAGCTCGGATGTTCTG 988
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 989 AATGACTTTCTTAATGATCAAAACAGAGATAAACAGATATTTACAGTGAAGTGAAGCC 1048
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 1049 TTGGGGACGTTAAATTAGAAAGATTACTTACTTTGACAGAGATGTGAAGGACATTTATG 1108
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 1109 AAGGTAGTTCCTGAGCTATTCAACACAGAAACAAAGCGCTTTGCTGTGCTAATCAG 1168
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1169 AATGGAATAATGTACACATCCTCTGTCTATGTGTTCCCTGGCTTCTGTCTGGCACAGTAC 1228

QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1229 TCACCTCAGCAGTTGGCAGGGAAGAGGTTGAGTGTCTCTTACGGTTCTGGCTTGCT 1288
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1289 GCCACGCTGTACTCCCTTAAAGTCACACAAAGCGCCACACAGATCTGCCCTGTGATAA 1348
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1349 ATAACAGCAAGTTATGTGACCTTAAATCAAGGCTGACTCGAAGAACTGTGTGGACCG 1408
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db 1409 GATGCTTTGCTGAAAAACATGAAGCTCAGAGAGGACACACATCACTTAGCCAACTATATT 1468
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
Db 1469 CCCAGTGTCAATAGACTCACTCTTTGAAGGAACGTGTATCTGTCAGAGTGAAGAA 1528
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1529 AAACACAGAAGACTTACGCGCGCGCCCTTCACAAATGACCACAGTTTGATGAAGAA 1588
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db 1589 ATGGGGCTCGTCATAGTACACAGCAACAGAGCATATTCCAAGCCCTGCTAAGAAAGTG 1648
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaValIleSerAsnGlyGluHis 478
Db 1649 CCAAGACTCCCTGCAACCTCGCGCCGAATCTGAATCAGCTGTCACTAAGTAAACGGGAGCAC 1708
RESULT 11
BC034317
LOCUS 3202 bp mRNA linear ROD 12-NOV-2003
DEFINITION Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
(CDNA clone MGC:36662 IMAGE:5366786), complete cds.
ACCESSION BC034317
VERSION BC034317.1 GI:21706865
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3202)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heieh,F.,
Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahney,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 3202)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer

REMARK
COMMENT

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

FEATURES
source
1. .3202
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="MGC:36662 IMAGE:5366786"
/tissue_type="Eye, retina, mouse strain C57Bl\6"
/clone_id="NIH_MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .3202
/gene="Hmgc61"
/note="synonym: MGC36662"
/db_xref="LocusID:208715"
/db_xref="MGI:107592"
25. .1587
/codon_start=1
/product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1"

CDS
/protein_id="AAH34317.1"
/db_xref="GI:21706866"
/db_xref="LocusID:208715"
/translation="MPGSLPLNAEACWPKDVGIVALEIYFPQYVDQAELEKYDGVDA
GKYTIGLQARMGFCTDREDINSCLTVQKLMERHSLSYDCIGRLVETETIIDKSK
SVKSNIMQLFEEGNTDIEGIDTINACYGTAVPNAVNVWESSWDGRYALVAGDI
AIYATGNARPTGGVAVALLIGPAFLIFDRGLGTHMOHAYDFYKPDMLSEYVVDG
KLSIQCYLSALDRCSVYRKIKTAQWQKEGKDFTLNDFGFMTPHSPYCKLVOKSLA
RMFLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMAKASELFNQKTKAS
LIVSNONGMYTSSVYGSILASVLAQYSPQQLAGKRVFVSYSGSLAATLYSLKVTQDA
TPGSALDKITASI CDLKSRLDSRTCVAPDVFAENMKLREDTHLANIYIPQCSIDSLFE
GTWYLVRVDEKHKRRTYARRPFNDHSLDEGMGLVHSNTATEHTIPSPAKKVPRLPATSA
ESESAVISNGEH"
61. .1431
/note="HMG-CoA synt; Region:
Hydroxymethylglutaryl-coenzyme A synthase"
/db_xref="CDD:pfam01154"

misc_feature
61. .1431
/note="HMG-CoA synt; Region:
Hydroxymethylglutaryl-coenzyme A synthase"
/db_xref="CDD:pfam01154"

ORIGIN

Alignment Scores:
Pred. No.: 7.29e-191 Length: 3202
Score: 2348.00 Matches: 452
Percent Similarity: 89.42% Conservative: 13
Best Local Similarity: 86.92% Mismatches: 13
Query Match: 93.51% Indels: 42
DB: 10 Gaps: 1

US-10-622-516-2 (1-478) x BC034317 (1-3202)

QY 1 MetProGlySerLeuProLeuAAsnAlaGluAlaCysTrpProLyAspValGlyIleVal 20
Db 25 ATGCCTGGGTCACTTCCTTGAATGCAAGGCTTGCTGCCCAAAAGATGTGGAATCCTT 84

QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluIleTyrAsp 40
Db 85 GCCCTTGAATCTACTTCTCTCTCATATATGTCATCAAGCTGAGTTGAAAAATACGAT 144
QY 41 GlyValAspAlaGlyIleTyrThrIleGlyLeuGlyGlnAlaIleMetGlyPheCysThr 60
Db 145 GGTGTAGATGCTGGAAAGATATACCATCGGCTGGGCCAGGCCAGGATGGCTTCTGCACG 204
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn 80
Db 205 GATCGTGAAGACATCAACTCTCTTGGCTGACTGTGTTCAGAAACTGATGAGAGACAT 264
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 265 AGCCTTCTATGATTGCAATTGGCGCGCTAGAAATTGGAACAGACAAATCATCGACAAA 324
QY 101 SerLysSerValIleThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 325 TCGAAATCAGTGAAGTCTAAATTGATGCAGCTGTTGAGGAGTCTGGGAATACAGATATA 384
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 385 GAAGGAATAGATACAAACCAATGCATGTGCGGGCCACAGCTGCAGTCTTCAATGCCGTG 444
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 445 AACTGGGTGAATCCAGCTCTTGGATGACGATATGCTCTGTAAGTTCAGAGACATATT 504
QY 149 ----- 149
Db 505 GCTATATATGCCACAGAAATGCCAGACCTACAGGTGAGTTGAGCTGTGGCCCTGCTA 564
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 565 ATTGGGCCAAACGCTCCTTAATTTTGAACGAGGCTCGGTGGAGACACATGACAGCAT 624
QY 159 AlaTyrAspPheTyrIleProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 625 GCCTATGACTTTTACAAAGCTGACATGCTCTCCAGTACCTGTGTGTCGACGAAAGCTC 684
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIle 198
Db 685 TCCATACAGTGTCACTCAAGCGCCCTGGACCGCTGCTATTCTGTCTACCGCAAAAAGATC 744
QY 199 HisAlaGlnTrpGlnIleGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 745 CGTGCCAGTGGCAGAAAGAGGAAAGCATTAAGATTTAACCTGAATGATTTTGGCTTC 804
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnIleSerLeuAlaArgMetLeuLeu 238
Db 805 ATGATCTTCACTACCATATTTGTAACGTGTGCAGAAATCTTAGCTCGGATGTTCCCTG 864
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 865 AATGACTTCTTAATGATCAAAAACAGAGATAAAAACAGTATTTACAGTGAAGCTGAGACC 924
QY 259 PheGlyAspValIleLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 925 TTGGGAGCGTTAAATTAGAAAGATCTTACTTGCACAGAGATGTAGAAAAGCATTTATG 984
QY 279 LysAlaSerSerGluLeuPheSerGlnIleTyrThrLysAlaSerLeuLeuValSerAsnGln 298
Db 985 AAGGCTAGTCTGAGCTATTCAACACAGAAACAAAGCGCTTGTGCTGCTAATCAG 1044
QY 299 AsnGlyAsnMetIleTyrThrSerSerValTyrGlySerLeuAlaSerValleuAlaGlnTyr 318
Db 1045 AATGGAATATATGACACATCCTGTCTATGTTGCTTCCCTGCTTGTGCGACAGTAC 1104
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1105 TCACCTCAGCAGTGGCCAGGAGAGGGGTTGAGTGTCTCTTACGGTTCTGCGCTTGCT 1164

QY	339	AlaThrLeuTyrSerLeuLysValThrglnAspAlaThrProGlySerAlaLeuAspLys	358
DB	1165	GCCACGCTGTACTCCCTTAAGTCACACAGACGCCACACGAGATCTGCCCTTGATAA	1224
QY	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrglyValAlaPro	378
DB	1225	ATAACAGCAGTTTATGTGACCTTAATCAAGGCTTGACTCGAGAACTTGTTGGCACCG	1284
QY	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle	398
DB	1285	GATGCTTTGGCTGAAAACATGAAGCTCAGAGAGGACACATCACTTAGCCCACTATATT	1344
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu	418
DB	1345	CCCCAGTGTTCATAGACTCACTCTTGAGGAACGTGATCTGTCAGAGTGGATGAA	1404
QY	419	LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly	438
DB	1405	AAACACAGAGAGACTTACGCCCGCGCCCTTCACAATATGACCACAGATTGGATGAAGA	1464
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysIleVal	458
DB	1465	ATGGGGCTCGTGCACTAGTAACAACAGCACAGAGCATATTCCAAAGCCCTGCTAAGAAAGTG	1524
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
DB	1525	CCAAAGACTCCCTGCACACTCGGCCGAATCTGAATCAGCTGTCACTAAGCGGGAGAC	1584

RESULT 12	BC029693	LOCUS	BC029693	3260 bp	mRNA	linear	ROD 07-OCT-2003
DEFINITION	Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase A (cDNA clone MGC:36525 IMAGE:5375374), complete cds.						

ACCESSION	BC029693	GI:20988708
VERSION	BC029693.1	
KEYWORDS	MGC.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	

REFERENCE
AUTHORS
1 (bases 1 to 3260)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Igoellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE	22388257
PUBMED	12477932
REFERENCE	2 (bases 1 to 3260)
AUTHORS	Strausberg, R.

TITLE Direct Submission
JOURNAL Submitted (06-MAY-2002) National Institutes of Health, Mammalian

REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk

Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunararatne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisseged, H.,
Kowls, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanaavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 53 Row: p Column: 14
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22122370.

FEATURES	Location/Qualifiers
source	1. .3260

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="MGC:36525 IMAGE:5375374"
/tissue_type="Mammary tumor. MM1V-LTR/INT3 model. 5 month
old mouse. Taken by biopsy."
/clone_lib="NCI CGAP_Mam2"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..3260
gene

```

```

/gene="Hmgcs1"
/note="synonym: MGC36662"
/db_xref="LocusID:208715"
/db_xref="MGI:107592"
79.1641
CDS

```

```
misc_feature
115..1485
/note="HMG CoA synt; Region:
Hydroxymethylglutaryl-coenzyme A synthase"
/db_xref="CDD:pfam01154"
/codon_start=1
/product="Hmgcs1 protein"
/protein_id="AAH29693.1"
/db_xref="GI:20988709"
/db_xref="locusID:208715"
/translation="MPGSLPLNAEACWPKDVGIVALEIYFSPQYVDQAELEKYDGVDA
GYKTITGLQARMGFCITDREDINSLCLTVVQKMERHSLSYDCIGRLEVGTEETIIDKSK
SVKSNLMQLFEESGNTDIEGIDTINACYGCGTAAVENAVNWVESSSWDGRYALVAGDDI
AIVATGNARPTGVGAVALLIGENAPLI FDRGLRGTHMQAHYDFYKPDMLSEYFVVDGQ
KLISIQCYLSALDRCYSVYRKKIRAQWQKEGKDKDFTLNDFGFMI FHSPYCKLVQKSLA
RMFLNDFLNDQNRDKNSIYSGLEAFGDVXLEDITYPRDVEKAFMKASSELFNOKTKAS
LTVSNQNGNMYTSSVYGSLSLADSLTAQVSPQOLAGKRVFYSYSGLAATLYSLKTYDDA
TPGSLNDKXITASLCLDKSRSLDSRTCAVPDVAEENKLRBDTHLANYI PQCSIDSLFEE
GTWTLVLRVDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATTEHI PSPAKKVPRLPATSA
ESESAAVINSNGEH"
```

ORIGIN

Alignment Scores:	
Pred. No.:	7.47e-191
Score:	2348.00
Percent Similarity:	89.42%
Best Local Similarity:	86.92%
Query Match:	93.51%
DB:	10
	Gaps: 1
	Length: 3260
	Matches: 452
	Conservative: 13
	Mismatches: 13
	Indels: 42
	Gaps: 1

US-10-622-516-2 (1-478) x BC029693 (1-3260)

```

QY      1 MetProGlySerIeuProLeuAsnAlaGluAlaCysTrpProLyAspValGlyIleVal 20
          |||||
DB      79 ATGCTGGGTCACTTCCTTTGAATGCAGAGGCTTGTGCGCCAAAGATGTGGGAATCGTT 138

```

21 A l a l e u g l u i l l e r y p h e p r o s e r g i n t y r v a l a s p g l n a l a g l u l e u g l u l y s t y r a s p 400

Db 139 GCCCTGAATCTACTTCCCTTCTCAATATGTCGATCAAGCTGAGTTGGAATAATACGAT 198
Qy 41 G1yValAspAlaG1yLysTyrThrIleG1yLeuG1yGlnAlaLysMetG1yPheCysThr 60
Db 199 GGTGTAGATGCTGGAAGATATACCATCGGCTGGGCCAGGCCAGGATGGGCTTCTGCACG 258
Qy 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 259 GATCGTAGAGACATCAACTCTCTTGCCCTGACTGTGTTCAAGAACTGATGAGAGACAT 318
Qy 81 AsnLeuSerTyrAspCysIleG1yArgLeuGluValG1yThrGluThrIleIleAspLys 100
Db 319 AGCCTTTCCTATGATTGCATTGGGCGGCTAGAGATTGGAACAGAGACAATCATCGACAAA 378
Qy 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerG1yAsnThrAspIle 120
Db 379 TCGAATACAGTGAAGTCTAATTGATGACAGCTGTTGAGGAGTCTGGAAATACAGATATA 438
Qy 121 GluG1yIleAspThrThrAsnAlaCysTyrG1yG1yThrAlaAlaValPheAsnAlaVal 140
Db 439 GAAAGAAATAGATACAACCAATGCATGCTATGGGGGCACAGCTGCAGTCTCAATGCCGTG 498
Qy 141 AsnTrpIleGluSerSerSerSerTrpAsp----- 149
Db 499 AACTGGGTCCAATCCAGCTCTTGGGATGAGCATATGCTGTGTAAGTTGCAGAGACATT 558
Qy 149 ----- 149
Db 559 GCTATATATGCCACAGAAATGCCAGACCTACAGGTGAGTTGAGAGCTGTGCCCTGCTA 618
Qy 150 -----GlyLeuArgG1yThrHisMetGlnHis 158
Db 619 ATTGGGCAAAAGCTCCTCTAATTTTGGACCGAGGGCTCCGTGGGACACACATGCAGCAT 678
Qy 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspG1yLysLeu 178
Db 679 GCCTATGACTTTTACAAGCCTGACATGCTCTCCGAGTACCTGTGTGTCGACGAAAGCTC 738
Qy 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 739 TCCATACAGTGTCTACCTCAGCGCCCTGACCGCTGCTATTGTCTACCGCAAAAAGATC 798
Qy 199 HisAlaGlnTrpGlnLysGluG1yAsnAspLysAspPheThrLeuAsnAspPheG1yPhe 218
Db 799 CGTGCCAGTGGCAGAAAGGGGAAAGGATTAAGATTTCCTGAATGATTTTGGCTTC 858
Qy 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 859 ATGATCTTTCACCTCACCAATATTGTAACGTGTGCAGAAATCTCTAGCTCGGATGTTCTCTG 918
Qy 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerG1yLeuGluAla 258
Db 919 AATGACTTCTTAATGATCAAAAACAGAGATAAAAACAGTATTACAGTGCAGTGGAAAGCC 978
Qy 259 PheG1yAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 979 TTTGGGAGCTTAAATTAGAGATACTTACTTTGACAGAGATGTAGAAAGGCATTATG 1038
Qy 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 1039 AAGGCTAGTCTGAGCTATTCACAACGAGAAACAAAGGCGCTTTGCTTGCTTAATCAG 1098
Qy 299 AsnG1yAsnMetTyrThrSerSerValTyrG1ySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1099 AATGAAATATGTACACATCCTCTGTCTATGGTTCCCTGGCTTCTGCTGCACAGTAC 1158
Qy 319 SerProGlnGlnLeuAlaG1yLysArgIleG1yValPheSerTyrG1ySerG1yLeuAla 338
Db 1159 TCACCTCAGCAGTTGGCAGGGAAGAGGTTGAGTGTCTCTTACGGTTCTGGCTTGCT 1218
Qy 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProG1ySerAlaLeuAspLys 358
Db 1219 GCCACGCTGTACTCCCTTAAGTACACACAGACGCCACACAGGATCTGCCCTTGATAA 1278

Qy 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrG1yValAlaPro 378
Db 1279 ATAACAGCAAGTTTATGTGACCTTAAATCAAGGCTTGACTCGAGAACTGTGTGGCACCG 1338
Qy 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle 398
Db 1339 GATGCTTTGCTGAAAAACATGAAGCTCAGAGAGACACACATCACTTAGCCAACTATATT 1398
Qy 399 ProGlnG1ySerIleAspSerLeuPheGluG1yThrTrpTyrLeuValArgValAspGlu 418
Db 1399 CCCAGTGTTCATATGACTCACTCTTTGAAGGAACGTGTATCTGTCAAGAGTGAATGAA 1458
Qy 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1459 AAACACAGAGGACTTACGCCCGCGGCCCTTCACAAAATGACCACAGATTGGATGAAGA 1518
Qy 439 ValG1yLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db 1519 ATGGGGCTCGTGCAATAGTAACACAGCAACAGAGCATATTCCAAGCCCTGCTAAGAAAGTG 1578
Qy 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnG1yGluHis 478
Db 1579 CCAAGACTCCCTGCAACCTCGGCCGAATCTGAATCAGCTGTCAATCAGTAACGGGAGACAC 1638
RESULT 13
BC031363 3278 bp mRNA linear ROD 07-OCT-2003
LOCUS
DEFINITION Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
(CDNA clone MGC:36633 IMAGE:5357048), complete cds.
ACCESSION BC031363
VERSION BC031363.1 GI:21618632
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 3278)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McSwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 3278)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Offices, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.

Qy 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db 1341 GATGCTCTTGGTGAACATGAGCTCAGAGAGACACATCACTTAGCCAACTATATT 1400
Qy 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
Db 1401 CCCAGTGTTCATAGACTCACTCTTTGAAGAACGTGTATCTGTGCAGAGTGAAGAA 1460
Qy 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1461 AAACACAGAAAGACTTACGCCCGCGCCCTTCACAAATGACCAAGTTTGGATGAAGGA 1520
Qy 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db 1521 ATGGGGCTCGTGCATAGTAACACACAGACATATTCCAAGCCCTGCTAAGAAAGTG 1580
Qy 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db 1581 CCAAGACTCCCTGCACCTCGGCCGAATCTGAATCACTGTCAATCACTAACGGGGAGCAC 1640

RESULT 14
CHKMGCOAS 1874 bp mRNA linear VRT 06-MAR-1995
LOCUS Chicken HMG-CoA synthase mRNA, complete cds.
DEFINITION M60657.1 GI:211930
VERSION 3-hydroxy-3-methylglutaryl-CoA synthase; HMG-CoA synthase.
KEYWORDS Gallus gallus (chicken)
SOURCE Gallus gallus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 1874)
AUTHORS Kattar-Coolley, P.A., Wang, H.H., Wende-Mueller, L.M. and Miziorko, H.M.
TITLE Avian liver 3-hydroxy-3-methylglutaryl-CoA synthase: distinct genes encode the cholesterologenic and ketogenic isozymes
JOURNAL Arch. Biochem. Biophys. 283 (2), 523-529 (1990)
MEDLINE 91112772
PUBMED 1980405

COMMENT Original source text: Chicken liver, cdNA to mRNA, clone NC-9.
FEATURES
source location/Qualifiers
1..1874
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="NC-9"
/tissue_type="liver"
/tissue_lib="Clontech"
1..1874
/gene="HMG-CoA synthase"
50..1618
/gene="HMG-CoA synthase"
/codon_start=1
/product="3-hydroxy-3-methylglutaryl-CoA synthase"
/protein_id="AAA62737.1"
/db_xref="GI:211931"

gene
CDS

/translation="MPGSLPVNTESCMPKDVGIValEiYFPsQYVDQTELEKXDEVDA
GKTTIGLGQSKMGFCSDREDINSLCLTVQKIMERNLSYDCTGRLEVGTEITIIDSKS
SVKTVLMOLFEESENTDVEGIDITNACYGTALFNAINWIESSWDGRYALVAGDI
AVYATGNARPTGGAGAVAMVGSNAPLIFERGLRGTHMQHAYDFYKPDMSVSEYVDG
KLSIQCYLSALDRCSVYRNKTHAQWQKEGTDRGFTLNDFGFMI FHS PYCKLVOKSVA
RLIINDFLSDQNAETANGVFGSLAEAFRDVKLEDYEDRDVEKAFMKASAELEFNQTKA
SLVSNQNGNMWTPSVYGLASILAQYSPEHLAQRISEFSYSGFAATLYSIRVTQD
ATPGSALDKITAGISDLKARLDSKRCIAPDVFAENMKIROETHLIANYIPOCSYEDLF
EGTWYLVVRVDEKHKRTYARRPVMGDGPLAEAVEVHPGIVHEHIFSPAKKVPRIPAT
ESEGVTVAIISNGVH"

ORIGIN

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:

1.04e-169 length: 1874
2097.00 Matches: 400
83.52% Conservative: 36

Best Local Similarity: 76.63% Mismatches: 42
Query Match: 83.51% Indels: 44
DB: 5 Gaps: 3
US-10-622-516-2 (1-478) x CHKMGCOAS (1-1874)

Qy 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 50 ATGCCCGGCTCTTCCAGTGAACACTGAATCCTGCTGGCCCAAGATGTGGTATTGTT 109
Qy 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 110 GCACTGGAATCTAATTTTCCCTTCAGTATGTGACCAAGACTGAGCTGGAAGATGAC 169
Qy 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 170 GGTGTGATGCAGGCAAAATACACCAATTGGGTTAGGCCAGTCMAAGATGGGCTTCTGCT 229
Qy 61 AspArgGluAspIleAsnSerLeuCysMetThrValAlaGlnAsnLeuMetGluArgAsn 80
Db 230 GACCGAAGAGATATCAATTCCTCTGTGTTGACTGTCTCAGAAGCTTATGAGAGAAC 289
Qy 81 AsnLeuSerTyrAspCysIleGlyValArgLeuGluValGlyThrGluThrIleLeuAspLys 100
Db 290 AGCCTTTCCTATGACTGCATTGGGAGACTGGAAGTTGGAACGAGACAAATTAATTGATTA 349
Qy 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 350 TCAAAATCGGTGAAGACTGTCTGTGACAGCTATTGTAAGAATCTGTAAATACAGATGTA 409
Qy 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 410 GAAGGAATTGACACACCAATGCGTGTATGGAGGCACTGCTGCTTTTAATGCTATT 469
Qy 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 470 AACTGGAATTGAGTCCAGTTCTTGGGATGAGACGCTATGCACCTTGTGTTGCTGAGACATT 529
Qy 149 ----- 149
Db 530 GCTGTATATGCCACTGGAAATGCCAGGCCAACAGGTGAGACTGTGCTGTATGCTA 589
Qy 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 590 GTTGGGTCAATGCTCCTTTAATTTTGAAGAGAGATTGCGTGAACCCACATGACAGCAT 649
Qy 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 650 GCTTATGACTTCTATAACCAAGATATGTTCTGAATATCTGTAGTGAAGCAAACTA 709
Qy 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIle 198
Db 710 TCTATACAGTGTCACTACCTGATGACATGACCGCTGTACTAGTGTATTCGAATAAATC 769
Qy 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 770 CATGCCCAGTGGCAAAAAGAGGGGACAGACAGAGGTTTCACCTTGAATGATTTGGAATC 829
Qy 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 830 ATGATCTTCAATTCTCCCTACTGTAAACTGTGAAGAGTGGTGGCAAGACTGTTGCTG 889
Qy 239 AsnAspPheLeuAsnAspGlnAsnArgAsp---LysAsnSerIleTyrSerGlyLeuGlu 257
Db 890 AATGACTTCTCAGTGACCAAGATGCAGAAACAGCAAAATGTTGTTTCACTGCTCGAA 949
Qy 258 AlaPheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPhe 277
Db 950 GCTTTCAGGAGATGTAAGCTTGAAGATACATATTTTGATAGGATGTGAAAAGCTTTT 1009
Qy 278 MetLysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsn 297
Db 1010 ATGAAGCTAGTGACAGAGCTTTCATCAATCAGAAAACCAAGCTTCTTACTGTGTCCAAT 1069

Qy	298	GlnaSnglYasnMctYrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGln	317
Db	1070	CAGAATGGAAACATGTACACGCCCTTCACTACGGTGCCTTCTCTTACGCCAG	1129
Qy	318	TyrSerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeu	337
Db	1130	TACTCTCCAGACACCTTCGACGACAAGAATCAGTAGTTCATATGCTCTGTTT	1189
Qy	338	AlaAlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAsp	357
Db	1190	GCTGCTACGCTGATTCATCAGAGTTACACAGGATGCCATCCTGTTCTGCGCTTGAC	1249
Qy	358	LysIleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAla	377
Db	1250	AAATAACTGCTAGCCTTCTCTGATCTTAAGCAAGACTTGACTCAGAAAATGCATTGCA	1309
Qy	378	ProAspValPheAlaGlnAsnMctLysLeuArgGluAspThrHisIleLeuValAsnTyr	397
Db	1310	CCTGATGCTTGTGCGAAACATGAAGATTAGACAGAGACACATCACTTGCCAACTAT	1369
Qy	398	IleProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAsp	417
Db	1370	ATTCACAGTGTTCACTAGAGACATCTCTTGAGGGAACATGTAATCTTGCTGCTGGAT	1429
Qy	418	GluLysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGlu	437
Db	1430	GAAAAACACAGAGAACATATGCACGACGCCAGTTATGGGTGATGACCCCTGGAGCA	1489
Qy	438	GlyValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLys	457
Db	1490	GGAGTTGAAGTTGTCCACCCAGCATTTGTCATGAGCACATCCCAAGCCCTGTAAGAA	1549
Qy	458	ValProArgLeuProAlaThrAla--AlaGluProGluAlaAlaValIleSerAsnGly	476
Db	1550	GTGCCAAGAATCCCTGCACACACAGAATCTGAAGGCGTTACTGTGCCATTCCAAATGG	1609
Qy	477	GluHis	478
Db	1610	GTGCAT	1615
RESULT	15		
LOCUS	165510	165510	1824 bp
DEFINITION	Sequence 1 from patent US 5668001.	DNA	linear
ACCESSION	165510		PAT 07-OCT-1997
VERSION	165510.1	GI:2482080	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 1824)		
TITLE	Mizioro, H.M.		
JOURNAL	3-hydroxy-3-methyl-glutaryl-coa synthase preparation with improved		
FEATURES	stability		
source	Patent: US 5668001-A 1 16-SEP-1997;		
	Location/Qualifiers		
	1..1824		
	/organism="unknown"		
	/mol_type="unassigned DNA"		
ORIGIN			
Alignment Scores:			
Pred. No.:	4.85e-169	length:	1824
Score:	2089.00	Matches:	399
Percent Similarity:	83.33%	Conservative:	36
Best Local Similarity:	76.44%	Mismatches:	43
Query Match:	83.19%	Indels:	44
DB:	6	Gaps:	3
US-10-622-516-2 (1-478) x 165510 (1-1824)			
Qy	1	MctProGlySerLeuProLeuAsnAlaGlyAlaCysTrpProLysAspValGlyIleVal	20

Db	2	ATGGCTGGGTCTCTCCAGTGAACACTGAATCCCTGCTGGCCCAAGATGTGGTATGT	61
Qy	21	AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluIleTyrAsp	40
Db	62	GCACTGGAATCTATTTCCTCTCAGTATGTCGACCAGACTGAGCTGGAAGTATGAC	121
Qy	41	GlyValAspAlaGlySerTyrThrIleGlyLeuGlyGlnAlaIleMetGlyPheCysThr	60
Db	122	GGTGTGATGACGGCAATAACACCATTTGGTTAGGCCAGTCAAAAGATGGCTTCGTCT	181
Qy	61	AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn	80
Db	182	GACCGAGAGATATCAATCCCTCTGTTGACTGTCGTTCAAGACTTATGAGAGAGAAC	241
Qy	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspIle	100
Db	242	AGCCTTCTATGACTGATGCTGGGAGACTGGAAAGTTGGAAACGAGACAATTAATTGATAA	301
Qy	101	SerIysSerValIysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
Db	302	TCAAAATCGGTGAAGACTGTCCTGATGACAGCTATTGTGAAGAATCTGTAATACAGATGA	361
Qy	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
Db	362	GAAAGAATTGACACACCAATCGCTGATGAGAGGCACTGCTCTTTTAATGCTATT	421
Qy	141	AsnTrpIleGluSerSerSerTrpAsp-----	149
Db	422	AACTGATTGAGTCCAGTTCTTGGATGAGACGCTATGCACTTGTGTTGCTGAGACATT	481
Qy	149	-----	149
Db	482	GCTGTGATGCCACTGGAAATGCCAGGCCAACAGGTGAGCTGTGCTGTGCTATGCTA	541
Qy	150	-----GlyLeuArgGlyThrHisMetGlnHis	158
Db	542	GTTGGTCAATGCTCCTTTAATTTTGAGAGAGGATTGGCTGGAACCCACATGCAGCAT	601
Qy	159	AlaTyrAspPheTyrIysProAspMetLeuSerGluTyrProIleValAspGlyIysLeu	178
Db	602	GCTTATGACTTCTATTAACCAAGATATGTTTCTGAATATCCTGTAGTTGACCAACTA	661
Qy	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysIysIle	198
Db	662	TCTATACAGTGTACTCACTCACTGATTAGACCGCTGCTAAGTGTATTATCGCAATAAATC	721
Qy	199	HisAlaGlnTrpGlnIysGluGlyAsnAspIysAspPheThrLeuAsnAspPheGlyPhe	218
Db	722	CATGCCCACTGGCAAAAAGAGGGGACAGACAGAGGTTTCACTTGAATGATTTTGGATTC	781
Qy	219	MetIlePheHisSerProTyrCysIysLeuValGlnIysSerLeuAlaArgMetLeu	238
Db	782	ATGATCTTTCATTCTCCCTACTGTAACTGGTACAGAGTGGGTGGCAAGACTGTGCTG	841
Qy	239	AsnAspPheLeuAsnAspGlnAsnArgAsp--LysAsnSerIleTyrSerGlyLeuGlu	257
Db	842	AATGACTTCTCAGTGAACCAAGATGCAGAAACAGCAATGTGTTTTCAGTGTCTGAA	901
Qy	258	AlaPheGlyAspValIysLeuGluAspThrTyrPheAspArgAspValGluIysAlaPhe	277
Db	902	GCTTTCAGGATGTAAAGCTTGAAGATACATATTTGTATAGGGATGTGAAAAAGCTTTT	961
Qy	278	MetIysAlaSerSerGluLeuPheSerGlnIysThrIysAlaSerLeuLeuValSerAsn	297
Db	962	ATGAAAGCTAGTGCAGAGCTTTCATACAGAAAAACCAAGCTTCTTACTTGTGTCCAT	1021
Qy	298	GlnAsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGln	317
Db	1022	CAGAATGAAACATGTACACGCTTCACTACGTTGCTGCTTCTTCTTACGCCAG	1081
Qy	318	TyrSerProGlnIleuAlaGlyIysArgIleGlyValPheSerTyrGlySerGlyLeu	337

Db 1082 TACTCTCCAGAGCACTTGACAGACAAGAATCAGTAGTTCATATGCGCTCGGTTT 1141

Qy 338 AlaAlaThrLeuTyrSerLeuIysValThrGlnAspAlaThrProGlySerAlaLeuAsp 357
|||||
1142 GCTGCTACGCTTGTATTCATCAGAGTTACACAGGATGCCACTCCTGTTCTGCGCTTGAC 1201

Qy 358 LysIleThrAlaSerLeuCysAspLeuIysSerArgLeuAspSerArgThrGlyValAla 377
|||||
1202 AAAATTAAGCTAGCCTTCTGATCTTAAAGCAAGACTTGACTCAGGAAATGCATTGCA 1261

Qy 378 ProAspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyr 397
|||||
1262 CCGATGTCTTGTGTAAGAAACATGAGATTAGACAGAGACACATCACTTGCCCACTAT 1321

Qy 398 IleProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAsp 417
|||||
1322 ATCCACAGTGTTCAGTAGAAGATCTTTGAGGGAACATGTATCTTGTGCGTGGAT 1381

Qy 418 GluLysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGlu 437
|||||
1382 GAAAAACACAGAGAACATATGACAGACGCCAGTTATGGGTGATGGAACCCCTGAGGCA 1441

Qy 438 GlyValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLys 457
|||||
1442 GAGTTGAAGTTGTCCACCCAGGCAATTGTCATGAGCACATCCCAAGCCCTGCTAAGAAA 1501

Qy 458 ValProArgLeuProAlaThrAla--AlaGluProGluAlaAlaValIleSerAsnGly 476
|||||
1502 GTGCCAAGAAATCCCTGCACAACACAGAACTGAAGGCCGTTACTGTGCCATTTCCAATGGG 1561

Qy 477 GluHis 478
|||

Db 1562 GTGCAT 1567

Search completed: June 24, 2004, 15:10:38
Job time : 4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 24, 2004, 13:21:38 ; Search time 497 Seconds
(without alignments)
4085.792.Million cell updates/sec

Title: US-10-622-516-2

Perfect score: 2511
Sequence: 1 MPGLPLNNAEACWPKDVGIV.....PRLPATAAEPAAVISNGEH 478

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 segs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool_p/US10622516/runat_23062004_162632_639/app_query.fasta_1.647
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10622516 @CGN_1_1_708 @runat_23062004_162632_639 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	2511	100.0	2002	6	ABSS54409 Human hyd
2	2511	100.0	2002	7	ABX93299 CDNA enco
3	2480	98.8	1564	7	ACC62337 Human NOV
4	2480	98.8	1564	7	ACC62339 Human NOV
5	2480	98.8	1601	7	ACC62331 Human NOV
6	2480	98.8	1650	7	ACC62336 Human NOV
7	2480	98.8	1650	7	ACC62328 Human NOV
8	2480	98.8	1650	7	ACC62326 Human NOV

9	2480	98.8	1650	7	ACC62330	Acc62330 Human NOV
10	2480	98.8	1650	7	ACC62338	Acc62338 Human NOV
11	2480	98.8	1650	7	ACC62334	Acc62334 Human NOV
12	2480	98.8	1650	7	ACC62327	Acc62327 Human NOV
13	2480	98.8	1650	7	ACC62332	Acc62332 Human NOV
14	2480	98.8	1650	9	ADE38394	Ad38394 Human pro
15	2480	98.8	3722	10	ADE76933	Ad76933 Human CDN
16	2475	98.6	1593	7	ACC62329	Acc62329 Human NOV
17	2475	98.6	1608	7	ACC62341	Acc62341 Human NOV
18	2475	98.6	1608	7	ACC62333	Acc62333 Human NOV
19	2475	98.6	1612	7	ACC62340	Acc62340 Human NOV
20	2475	98.6	1612	7	ACC62335	Acc62335 Human NOV
21	2439	97.1	1685	9	ADD14748	Add14748 Human src
22	2395.5	95.4	2540	5	AA584743	Aas84743 DNA encod
23	2362	94.1	3275	8	ACF25333	Acf25333 Rat cytos
24	2362	94.1	3275	9	ADB58295	Ad58295 Toxicity-
25	2362	94.1	3275	9	ADB52844	Ad52844 Primary r
26	2089	83.2	1824	2	AAT89089	Aat89089 Avian 3-h
27	1631	65.0	3008	4	AAH34834	Aah34834 Human col
28	1513	60.3	2058	6	ABL69551	Ab169551 Prostate
29	1513	60.3	2058	6	ABN97315	Abn97315 Gene #381
30	1513	60.3	2058	7	ACC50182	Acc50182 Breast ca
31	1513	60.3	2058	9	ADB75349	Ad75349 Prostate
32	1513	60.3	2088	4	AAK52002	Aak52002 Human pol
33	1513	60.3	2377	9	ADE53675	Ade53675 Human pro
34	1511.5	60.2	1994	4	AAH22426	Aah22426 Rat mitoc
35	1511.5	60.2	1994	7	ABT41772	Abt41772 Toxicity-
36	1511.5	60.2	1994	9	ADB57988	Ad57988 Toxicity-
37	1511.5	60.2	1994	9	ADB52471	Ad52471 Primary r
38	1510	60.1	2132	9	ADE09760	Ad09760 Novel DNA
39	1500	59.7	2053	4	AAK52986	Aak52986 Human pol
40	1497.5	59.6	2093	3	AAc98123	Aac98123 Human col
41	1337	53.2	1896	4	ABL07457	Ab107457 Drosophil
42	1337	53.2	1939	4	ABL18435	Ab118435 Drosophil
43	1337	53.2	4921	4	ABL07456	Ab107456 Drosophil
44	1337	53.2	6522	4	ABL18434	Ab118434 Drosophil
45	1336	53.2	1752	9	ADE07588	Ad07588 Novel cod

ALIGNMENTS

RESULT 1	ABSS54409	standard; CDNA; 2002 BP.
ID	ABSS54409	standard; CDNA; 2002 BP.
XX	AC	ABSS54409;
XX	DT	22-NOV-2002 (first entry)
XX	DE	Human hydroxymethylglutaryl-coenzyme A synthase protein cDNA.
XX	KW	Human; gene; ss; chromosome 5; hydroxymethylglutaryl-coenzyme A synthase;
KW	KW	HMG-CoA synthase; cholesterologenesis; therapeutic; diagnostic; genotype;
KW	KW	antibody; synthase; carcinoma.
XX	OS	Homo sapiens.
XX	XX	
FH	Key	Location/Qualifiers
FT	5'UTR	1..144
FT		/*tag= a
FT	CDS	145..1581
FT		/*tag= b
FT	variation	/product= "Hydroxymethylglutaryl-coenzyme A synthase"
FT		replace(717,T)
FT		/*tag= c
FT		/standard_name= "Single nucleotide polymorphism"
FT	3'UTR	1582..2002
FT		/*tag= d
PN	US6436692-B1.	
XX		
PD	20-AUG-2002.	
XX		

PF 29-MAR-2001; 2001US-00819993.
XX
PR 29-MAR-2001; 2001US-00819993.
XX
PA (APPL-) APPLERA CORP.
XX
PI Gong F, Yan C, Di Francesco V, Beasley EM;
XX
DR MPI; 2002-689940/74.
DR P-PSDB; ABG32726.
XX
PT New isolated nucleic acid molecule encoding hydroxymethylglutaryl-CoA
PT synthase, useful as model for the development of human therapeutic
PT targets and for identifying therapeutic proteins.
XX
PS Claim 1; Fig 1; 62pp; English.
XX

CC The invention discloses an isolated nucleic acid molecule encoding a
CC human hydroxymethylglutaryl-coenzyme A (HMG-CoA) synthase which is
CC important in cholesterologenesis. The polynucleotide and polypeptide are
CC useful as models for the development of human therapeutic targets, to aid
CC in the identification of therapeutic proteins and as targets for the
CC development of human therapeutic agents that modulate the activity of the
CC polypeptide in cells and tissues. The polynucleotide is useful for
CC monitoring the effectiveness of modulating compounds on the expression or
CC activity of the enzyme gene in clinical trials and in treatments, in
CC diagnostic assays for qualitative changes in expression of enzyme nucleic
CC acid, to detect mutations in enzyme genes and gene expression products,
CC such a mRNA, for testing an individual for a genotype and as a diagnostic
CC target that can be used to tailor treatment in an individual. The
CC polypeptide is useful to raise antibodies, as a reagent in assays
CC designed to quantitatively determine levels of the protein in biological
CC fluids, as markers for tissues in which the corresponding protein is
CC preferentially expressed (e.g. carcinomas), in biological assays related
CC to members of the synthase subfamily, in drug screening assays, in
CC competition binding assays, in cell-based or cell-free systems, in
CC pharmacogenomic analysis and for treating a disorder characterised by an
CC absence of, inappropriate, or unwanted expression of the polypeptide. The
CC sequence presented is the human HMG-CoA synthase protein cDNA, the gene
CC for which is located on chromosome 5
XX

SQ Sequence 2002 BP; 590 A; 380 C; 472 G; 560 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.72e-253 Length: 2002
Score: 2511.00 Matches: 478
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-622-516-2 (1-478) x ABS54409 (1-2002)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db |||||
Db 145 ATGCCTGATCACTTCCTTGAATGCAGAGCTTGCTGCCAAAAGATGGGAATTGTT 204
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db |||||
Db 205 GCCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCAGAGCTGGAAAAATATGAT 264
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db |||||
Db 265 GGTGTAGATGCTGAAAGATATACCATGGCTGGGCCAGGCCAAGATGGGCTTCTGCACA 324
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db |||||
Db 325 GATAGAGAAGATATTACTCTCTTTGCATGACTGIGGTTCAGAATCTTATGAGAGAAAT 384
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db |||||
Db 385 AACCTTCTCTATGATGATTTGGCGGCTGGAAGTTGGAACAGACAGACAAATCATCGACAAA 444

QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db |||||
Db 445 TCAAAGTCTGTGAAGACTAATTGTATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA 504
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db |||||
Db 505 GAAGGAATCGACACAATAATGATGCTATGAGGACACAGCTGCTGCTTCAATGCTGCTT 564
QY 141 AsnTrpIleGluSerSerSerTyrPaspGlyLeuArgGlyThrHisMetGlnHisAlaTyr 160
Db |||||
Db 565 AACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGTGGGACACATATGCACAATGCCCTAT 624
QY 161 AspPheTyrLysProAspMetLeuSerGluTyrTrpProIleValAspGlyLysLeuSerIle 180
Db |||||
Db 625 GATTTTACAAGCCTGATATGCTATCTGAATATCCTATATAGTAGATGAAAACCTCTCCATA 684
QY 181 GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAla 200
Db |||||
Db 685 CAGTGTACCTCAGTGCATTTAGACCGCTGTACTCTGTCTACTGCCAAAAAGATCCATGCC 744
QY 201 GlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIle 220
Db |||||
Db 745 CAGTGCAGAAAGAGGGAATGATAAAGATTTTACCCTGAATGATTTGGCTTCATGATC 804
QY 221 PheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp 240
Db |||||
Db 805 TTTCACTCAACATATTGTAAACTGGTTCGAATCTCTAGCTCGGATGTTCCTGAATGAC 864
QY 241 PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGly 260
Db |||||
Db 865 TTCCTTAATGACCAGAAATAGAGATMAAATAATATATAGTGCGCTGGAAGCCTTTGGG 924
QY 261 AspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAla 280
Db |||||
Db 925 GATGTTAAATTAGAAAGACACTTACTTTGATAGAGATGGAGAAAGGCAATTATGAAGGCT 984
QY 281 SerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGly 300
Db |||||
Db 985 AGCTCTGAACCTTCAGTCAGAAACAAAGCATCTTTACTTGTATCAAAATCAAAATGGA 1044
QY 301 AsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerPro 320
Db |||||
Db 1045 AATATGTAACACATCTTCAATATATGTTCCCTTGCACTGTTCTAGCACAGTACTCACCT 1104
QY 321 GlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaIleThr 340
Db |||||
Db 1105 CAGCAATTAGCAGGGAAGAGAAATGGAGTGTTCCTTATGTTCTGGTTGGCTGCCACT 1164
QY 341 LeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThr 360
Db |||||
Db 1165 CTGTACTCTCTTAAGTCAACACAGATGCTACACCGGGGCTCTGCTTGTATAAATTAACA 1224
QY 361 AlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAspVal 380
Db |||||
Db 1225 GCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGACTGCTGGCACCATGTC 1284
QY 381 PheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIleProGln 400
Db |||||
Db 1285 TTCGCTGAAAAATGAAGCTCAGAGAGACACCATCATTTGGTCAACTATATATCCCCAG 1344
QY 401 GlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGluLysHis 420
Db |||||
Db 1345 GGTTCATATAGATCACTCTTTGAAGGAACGTGGTACTTAGTTAGGGTGATGAAAGACAC 1404
QY 421 ArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGlyValGly 440
Db |||||
Db 1405 AGAAGAAGCTTAGCTCGCGCTCCACCTCCAATGATGACACTTTGGATGAAGAGTAGGA 1464
QY 441 LeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysValProArg 460
Db |||||
Db 1465 CTGTGATTCATAAACAATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAGTAGCAAGA 1524
QY 461 LeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyLysHis 478

|||||
Db 1525 CTCCTGCCACAGACAGAACTGAAGCAGCTGTCTATTAGTAATGGGAACAT 1578
RESULT 2
ABX93299
ID ABX93299 standard; cDNA; 2002 BP.
XX
AC ABX93299;
XX
DT 05-JUN-2003 (first entry)
XX
DE cDNA encoding human HMG-CoA synthase-like enzyme.
XX
XX Human; cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase;
KM HMG-CoA synthase; immune response; drug screening assay;
KM pharmacogenomic analysis; chromosome 5; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH 5'UTR 1..144
FT /*tag= a
FT CDS 145..1581
FT /*tag= b
FT /product= "HMG-CoA synthase-like enzyme"
FT 3'UTR 1582..2002
FT /*tag= c
XX
PN US2002173018-A1.
XX
PD 21-NOV-2002.
XX
PF 12-JUL-2002; 2002US-00193295.
XX
XX 29-MAR-2001; 2001US-00819993.
XX
XX (APPL-) APPLERA CORP.
XX
PI Gong F, Yan C, Di Francesco V, Beasley EM;
XX
DR WPI; 2003-352594/33.
DR P-PSDB; ABU08379.
XX
XX
PT New isolated human synthase peptide and gene encoding the enzyme, useful
PT as models for developing human therapeutic targets, aid in the
PT identification of therapeutic proteins and as therapeutic targets.
XX
XX Claim 22; Fig 1A; 65pp; English.
PS
XX
CC The present invention relates to the isolation of a novel human enzyme
CC that shows a high degree of similarity to human cytoplasmic 3-hydroxy-3-
CC methylglutaryl coenzyme A (HMG-CoA) synthase. The invention also
CC discloses polynucleotide sequences encoding the novel enzyme of the
CC invention. Both the polypeptide and polynucleotide sequences are useful
CC as models for the development of human therapeutics, for identifying
CC therapeutic proteins, as targets for development of human therapeutic
CC agents, and as query sequences to perform a search against sequence data
CC bases to identify other family members of related sequences. The
CC polypeptide is useful to raise antibodies or to elicit another immune
CC response, as a reagent in assays designed to quantitatively determine
CC levels of the protein in biological fluids, as markers for tissues in
CC which the corresponding protein is preferentially expressed, in drug
CC screening assays, in cell-based or cell-free systems, to identify
CC compounds that modulate synthase activity of the protein in its natural
CC state, or an altered form that causes the specific disease or pathology
CC associated with the synthase, to screen a compound for the ability to
CC stimulate or inhibit interaction between the synthase protein and a
CC molecule that normally interacts with the synthase protein, and in
CC pharmacogenomic analysis. The polynucleotide is useful for monitoring the
CC effectiveness of modulating compounds on the expression or activity of
CC the human synthase gene in clinical trials or in a treatment regimen, in
CC diagnostic assays for qualitative changes in a human synthase nucleic
CC acid that leads to a pathology, for testing an individual for a genotype

CC that while not necessarily causing a disease, nevertheless affects the
CC treatment modality, and as antisense constructs to control human synthase
CC gene expression in cells, tissues and organisms. The present sequence
CC encodes human HMG-CoA synthase-like enzyme. The gene encoding the enzyme
CC is located on chromosome 5
XX
SQ Sequence 2002 BP; 590 A; 380 C; 472 G; 560 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 8.72e-253 Length: 2002
Score: 2511.00 Matches: 478
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0
US-10-622-516-2 (1-478) x ABX93299 (1-2002)
QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 145 ATGCGCTGATCACCCTCTTGAATGACAGAACTGTGGCCCAAAAGATGTGGAAATGTT 204
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 205 GCCCTGAGATCTATTTCCTCTCAATATGTTGATCAAGCAGAGTTGAAAAATATGAT 264
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 265 GGTGTAGATGCTGGAAGATATACCATTTGGCTTGGGCCAGGCCAAGATGGGCTTGCACA 324
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn 80
Db 325 GATTAGAAAGATATTAACTCTCTTGATGACTGTGTTCAGAAATCTTATGAGAGAAAT 384
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 385 AACCTTCTCTATGATTGATGGCGGCTGGAA GTGGAACAGAGCAATCATCGACAAA 444
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 445 TCAAAAGTCTGTAAGACTAATTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA 504
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 505 GAAGGAATCGACACACTAATGATGCTATGAGGCACAGCTGCTCTCAATGCTGTT 564
QY 141 AsnTrpIleGluSerSerSerTyrAspGlyLeuArgGlyThrHisMetGlnHisAlaTyr 160
Db 565 AACTGATTGAGTCCAGCTCTTGGGATGGGCTTCGTGGGACACATATGCAACATGCCCTAT 624
QY 161 AspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIle 180
Db 625 GATTTTACAAAGCTGATATGCTATCTGAATATCCTAATAGTAGTAGGAAAACTCTCATA 684
QY 181 GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIleHisAla 200
Db 685 CAGTGTACCTCAGTGCATTTAGACCCGCTGCTACTCTGTCTAATGCAAAAAGATCCATGCC 744
QY 201 GlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIle 220
Db 745 CAGTGGCAGAAAGAGGGAATATGAATTAAGATTTAACCTTGAATGATTTGGCTTCATGATC 804
QY 221 PheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp 240
Db 805 TTTCACCTACCATATTGTAACTGTTTCAGAAATCTCTAGCTCGGATGTTGCTGAATGAC 864
QY 241 PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGly 260
Db 865 TTCTTTAATGACCAGAAATAGATTAATAAATAGTATCTAATGTGGCCTGGAAGCCTTTGGG 924
QY 261 AspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAla 280
Db 925 GATGTTAAATTAGAAAGACACCTACTTGTATAGAGATGTGAGAAAGCATTTATGAAGGCT 984

QY	281	SerSerGluLeuPheSerGlnLysThrIleValAlaSerLeuLeuValSerAsnGlnAsnGly	300
Db	985	AGCTCTGAACCTCTTCAAGTCAGAAAAACAAGGCATCTTACTTGATCAAAATCAAAATGCA	1044
QY	301	AsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerPro	320
Db	1045	AATATGTACACATCTTCAGTATATGGTTCCTTGCACTGTTCTTAGCACAGTACTCACT	1104
QY	321	GlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThr	340
Db	1105	CAGCAATTAGCAGGGAGAGAATGGAGCTGTTTCTTATGGTTCGTGGCTGCCACT	1164
QY	341	LeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThr	360
Db	1165	CTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGCTCTGCTTGATAAATAACA	1224
QY	361	AlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAspVal	380
Db	1225	GCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGGTGGCACCAATGTC	1284
QY	381	PheAlaGluAsnMetLysLeuArgGluAspThrHisIleuValAsnTyrIleProGln	400
Db	1285	TTCGCTGAAAAACATGAAGCTCAGAGAGGACCCATCATTTGGTCAACATATATCCCCAG	1344
QY	401	GlySerIleAspSerLeuPheGluGlyThrTyrPtyrLeuValArgValAspGluLysHis	420
Db	1345	GGTTCATAGATTCACCTCTTGAAGGAACGTGGTACTTAGTGGGTGGATGAAGAAGCAC	1404
QY	421	ArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGlyValGly	440
Db	1405	AGAAGAACCTTACGCTCGCGCTCCCACTCCAATGATGACACTTGGATGAAGGAGTAGGA	1464
QY	441	LeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysValProArg	460
Db	1465	CTGTGCATTCAACATAGCACTGACCATATATCCAAAGCCCTGCCAAGAAAGTACCAAGA	1524
QY	461	LeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
Db	1525	CTCCCTGCCACAGACAGAACTGAAGACGCTGCATTAGTAATGGGGAACAT	1578
RESULT 3			
ACCG2337			
ID	ACCG2337	standard; cDNA; 1564 BP.	
XX	AC	ACCG2337;	
XX	DT	23-JUN-2003 (first entry)	
XX	DE	Human NOV411 encoding cDNA SEQ ID NO:203.	
XX	KW	Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological; anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfectility; haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator; neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeamic; gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma; congenital heart defect; aortic stenosis; valve disease; transplantation; tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes; prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer; fertility; haemophilia; hypercoagulation; graft versus host disease; idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia; Crohn's disease; multiple sclerosis; infectious disease; cancer; cancer-associated cachexia; Alzheimer's disease; Parkinson's disease; immune disorder; haematopoietic disorder; dyslipidaemia; metabolic syndrome X; gene; ss.	
OS	XX	Homo sapiens.	
XX	PN	WO2003023001-A2.	
XX	PD	20-MAR-2003.	
XX	PF	09-SEP-2002; 2002WO-US028538.	
XX	PR	07-SEP-2001; 2001US-0318120P.	
PR	07-SEP-2001; 2001US-0318184P.		
PR	10-SEP-2001; 2001US-0318430P.		
PR	17-SEP-2001; 2001US-0322636P.		
PR	17-SEP-2001; 2001US-0322781P.		
PR	17-SEP-2001; 2001US-0322816P.		
PR	17-SEP-2001; 2001US-0322817P.		
PR	19-SEP-2001; 2001US-0323519P.		
PR	20-SEP-2001; 2001US-0323631P.		
PR	20-SEP-2001; 2001US-0323636P.		
PR	25-SEP-2001; 2001US-0324969P.		
PR	25-SEP-2001; 2001US-0325091P.		
PR	26-SEP-2001; 2001US-0324990P.		
PR	14-DEC-2001; 2001US-0341144P.		
PR	26-FEB-2002; 2002US-0359599P.		
PR	05-MAR-2002; 2002US-0361663P.		
PR	03-MAY-2002; 2002US-0377908P.		
PR	17-MAY-2002; 2002US-0381483P.		
PR	29-MAY-2002; 2002US-0383863P.		
PR	02-JUL-2002; 2002US-0393332P.		
PR	17-JUL-2002; 2002US-0396412P.		
PR	13-AUG-2002; 2002US-0403517P.		
PR	06-SEP-2002; 2002US-00236417.		
XX	PA	(CURA-) CURAGEN CORP.	
PI	Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL, Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A, Crabtree J, DiIppio VA, Edinger SR, Eisen AJ, Ellerman K, Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W, Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X, Malyanekar UM, Miller CE, Oci CE, Ort T, Padigaru M, Patturajan M, Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;		
XX	WPI; 2003-313241/30.		
DR	P-PSDB; ABR54268.		
XX			
PT	Novel human proteins and nucleic acid encoding the proteins, useful for diagnosis, treatment and prevention of disorders involving the human protein or nucleic acid e.g. cardiac and neurological disorders.		
PT			
XX			
XX	Claim 20; Page 245-246; 460pp; English.		
XX			
CC	The present invention describes isolated human NOX proteins, where X is 1 to 42. ACCG2236 to ACCG2345 encode the human NOX proteins given in ABR54167 to ABR54276. NOX sequences have antiatherosclerotic, cardiant, hypotensive, dermatological, anorectic, immunosuppressive, cytostatic, antidiabetic, antiinfectility, haemostatic, antiinflammatory, anti-HIV, antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic, antiparkinsonian and antilipaeamic activities, and can be used in gene therapy. NOX proteins are useful for treating or preventing a pathology associated with a NOX protein in humans and for treating a syndrome associated with the human disease. NOX nucleic acids, proteins and antibodies can be used in the treatment and diagnosis of cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, valve disease, tuberosus sclerosis, scleroderma, obesity, transplantation, congenital adrenal hyperplasia, prostate cancer, fertility, haemophilia, disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis, infectious disease, anorexia, cancer-associated cachexia, cancer, Alzheimer's disease, Parkinson's disease, immune disorders, haematopoietic disorders, dyslipidaemias, and metabolic syndrome X. ACCG2346 to ACCG2465 represent PCR primers and probes for human NOX sequences, which are used in examples from the present invention. ABR54277 represents a human trypsinogen protein given in comparison with the human NOV35b protein in the exemplification of the present invention		
CC	Sequence 1564 BP; 464 A; 304 C; 362 G; 434 T; 0 U; 0 Other;		

Alignment Scores:

Pred. No.:	1.08e-249	Length:	1564
Score:	2480.00	Matches:	478
Percent Similarity:	91.92%	Conservative:	0
Best Local Similarity:	91.92%	Mismatches:	0
Query Match:	98.77%	Indels:	42
DB:	7	Gaps:	1

US-10-622-516-2 (1-478) x ACC62337 (1-1564)

QY	1	MetProGlySerLeuProLeuAenAlaGluAlaCysTrpProLysAspValGlyIleVal	20
DB	2	ATGCTTGATCACTCTCTTGATGCAGAAGCTTGCTGCCAAAGATGTGGGAATTGTT	61
QY	21	AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp	40
DB	62	GCCCTTGAGATCTATTCTCTCTCAATATGTGATCAAGCAGAGTTGAAAAATATGAT	121
QY	41	GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
DB	122	GGTGTAGATGCTGAAAGTATACCATTTGGCTTGGCCAGGCCAAGATGGCTTCTGCACA	181
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn	80
DB	182	GATAGAGAAGATATTAACTCTCTTGCACTAGCTGTGTTCAAGATCTTATGAGAGAAT	241
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
DB	242	AACCTTCTCATGATTCATTCATGGCGCGCTGGAAGTTGGAACAGACAATCATCGACAAA	301
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
DB	302	TCAAACTCTGTGAAGACTAATTTGATGCAGCTGTTGAAGAGTCTGGGAATACAGATATA	361
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
DB	362	GAAGGAATGCACACAACATAATGCATGTATGAGGACACAGCTGCTCTTCAATGCTGTT	421
QY	141	AsnTrpIleGluSerSerSerTrpAsp-----	149
DB	422	AACGTGATTGATCCAGCTCTTGGGATGACGGTATGCCCTGTAGTTGCAGAGATATT	481
QY	149	-----	149
DB	482	GCTGTATATGCCACAGGAATGCTAGACTACAGGTGAGTTGAGCAGTAGCTCTGCTA	541
QY	150	-----GlyLeuArgGlyThrHisMetGlnHis	158
DB	542	ATTGGGCCAAATGCTCCTTAATTTTGAACGAGGGCTTCGTGGACACATATGCAACAT	601
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu	178
DB	602	GCCATGATTTTACAGCCGTGATGCTATCTGAATATCTTAAGTAGATGAAAACTC	661
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
DB	662	TCCATACAGTGCTACCTCAGTGCAATTAGACCGCTGCTACTCTGCTACTGCAAAAAATC	721
QY	199	HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
DB	722	CATGCCCAATGGCAGAAAGAGGAATGATAAAGATTTTACCTTGATTTGGCTTC	781
QY	219	MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu	238
DB	782	ATGATCTTTCACTCACCATATTGTAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTG	841
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla	258
DB	842	AATGACTTCCTTAATGACCAGAAATAGATAAATAATAGTATCTATAGTGGCCTGGAAGCC	901
QY	259	PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet	278
DB	902	TTTGGGATGTAAATTAGAGACACCTACTTGTATAGAGATGTGGAAGGCAATTATG	961

QY	279	LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln	298
DB	962	AAGCTAGCTCTGAACCTCTTCACTCAGAAAAACAAGGCATCTTACTGTATCAATCAA	1021
QY	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
DB	1022	AATGAAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTCTAGCACAGTAC	1081
QY	319	SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
DB	1082	TCACTCAGCAATTAGCAGGGAAGAAATTGGAGTGTCTTATGTTCTGTGTTGGCT	1141
QY	339	AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys	358
DB	1142	GCCACTCTGTACTCTCTTAAGTCACACAAGATGCTACCGGGCTGCTCTTGATATAA	1201
QY	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
DB	1202	ATAACAGCAAGTTATGTGATCTTAATCAAGGCTTGATTCAGAAGCTGTGGCACCA	1261
QY	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle	398
DB	1262	GATGCTTCCTGTAACCATGAACTCTCAGAGAGACACCCATCATTTGTCACATATATT	1321
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu	418
DB	1322	CCCCAGGTTCAATAGATCACTCTTGAAGGAACGTGTACTTAGGTGATGATA	1381
QY	419	LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly	438
DB	1382	AAGCAGAGAAGACTTAGCGTCGGCGTCCCACTCCAATGATGACACTTTGATGAAGGA	1441
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal	458
DB	1442	GTAGACTTGTGCATTCAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA	1501
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
DB	1502	CCAAGACTCCCTGCCACAGCAGCAAACTGAAGCAGCTGTCAATGTAATGGGGAACAT	1561
RESULT 4			
ID	ACC62339	standard; cDNA; 1564 BP.	
XX	AC	ACC62339;	
XX	DT	23-JUN-2003 (first entry)	
XX	DE	Human NOV41n encoding cDNA SEQ ID NO:207.	
XX	KW	Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;	
KW	KW	anorectic; immunosuppressive; cyostatic; antidiabetic; antinfertility;	
KW	KW	haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;	
KW	KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaeimic;	
KW	KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;	
KW	KW	congenital heart defect; aortic stenosis; valve disease; transplantation;	
KW	KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;	
KW	KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;	
KW	KW	fertility; haemophilia; hypercoagulation; graft versus host disease;	
KW	KW	idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;	
KW	KW	Crohn's disease; multiple sclerosis; infectious disease; cancer;	
KW	KW	cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;	
KW	KW	immune disorder; haematopoietic disorder; dyslipidaemia;	
XX	XX	metabolic syndrome X; gene; ss.	
OS	OS	Homo sapiens.	
XX	XX	WO2003023001-A2.	
PN	PN	20-MAR-2003.	
XX	PD	09-SEP-2002; 2002WO-US028538.	
PF	PF		

XX 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Agee ML, Alsobrook JF, Anderson DW, Bergs C, Boldog FL,
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A,
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K,
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W,
PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X,
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patnrajan M,
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA,
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ,
PI Zernhusen BD, Zhong M,
XX
DR WPI; 2003-313241/30.
DR P-PSDB; ABR54270.
XX
PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS Claim 20; Page 246-247; 460pp; English.
XX

CC The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipemic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberosus sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1564 BP; 464 A; 304 C; 362 G; 434 T; 0 U; 0 Other;

Alignment Scores:			
Pred. No.:	1,08e-249	Length:	1564
Score:	2480.00	Matches:	478
Percent Similarity:	91.92%	Conservative:	0
Best Local Similarity:	91.92%	Mismatches:	0
Query Match:	98.77%	Indels:	42
DB:	7	Gaps:	1
US-10-622-516-2 (1-478) x ACC62339 (1-1564)			
QY	1	MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal	20
DB	2	ATGCCTGATCACTTCCTTGATGCAGAGCTTGCCCAAAAGATGGGAATTGTT	61
QY	21	AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp	40
DB	62	GCCCTTGAGATCTATTTCCTTCATATGTCAGCAGAGTTGAAATAATATGAT	121
QY	41	GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
DB	122	GGTGTAGATGCTGGAAGATATACCATTTGGCTGGCCAGGCCAAGATGGCTCTGCACA	181
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn	80
DB	182	GATAGAGAAGATATTAACTCTCTTGATGACTGTGTTCAAACTTATGAGAGAAAT	241
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleLeuAspLys	100
DB	242	AACCTTCTTATGATTGCAATTGGCGGCTGGAAAGTTGGAACAGACAAATCATGCACAA	301
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
DB	302	TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGCTCTGGGAATACAGATATA	361
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
DB	362	GAAGGAATGCACACAACTATGCATGTGAGGACACAGCTGCTGCTTCAATGCTGTT	421
QY	141	AsnTrpIleGluSerSerSerTrpAsp-----	149
DB	422	AACTGATGATGATCCAGCTCTTGATGACGCGTATGCCCTGTAATTGACAGAGATATT	481
QY	149	-----	149
DB	482	GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGACAGTAGCTCTGCTA	541
QY	150	-----GlyLeuArgGlyThrHisMetGlnHis	158
DB	542	ATTGGCCCAATGCTCTTAAATTTTGAACGAGGGCTTCGTGGACACATATGCAACAT	601
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu	178
DB	602	GCCATATGATTTTACAAAGCTGATATGCTATCTGAATATCTATAGTAGAATAAACTC	661
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
DB	662	TCCATACAGTGTCACTCACTGATGACATGACCGCTGCTACTCTGCTACTGCAGAAAAAGATC	721
QY	199	HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
DB	722	CATGCCAGTGGCAGAAAGAGGAAATGATTAAGATTTTAACTTGAATGATTTTGGCTTC	781
QY	219	MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu	238
DB	782	ATGATCTTCACTCACCATATTTAACTGTTTCAAGAACTCTAGCTGGATGTTGCTG	841
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla	258
DB	842	AATGACTTCTTAATGACCAAGATAGATATAAAATAGTATCTATAGTGCCCTGGAAGCC	901
QY	259	PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet	278
DB	902	TTTGGGATGTTAAATTAGAAAGACACCTACTTTGATAGAGATGTGAGAAAGCATTTATG	961

Qy	279	LYSAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuValSerAsnGln	298
Db	962	AAGGCTAGCTCTGAACCTCTTCACTCAGTACAGAAAAACAAGGCATCTTTACTGTATCAAAATCAA	1021
Qy	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
Db	1022	AATGAATAATGTACACATCTTCAGTATATGTTCCCTTGCACTGTCTAGACACAGTAC	1081
Qy	319	SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
Db	1082	TCACCTCAGCAATTAGCAGGAGAAGAATTGAGTGTITTTCTTATGTTCTGTGTTGGCT	1141
Qy	339	AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys	358
Db	1142	GCCACTCTGTACTCTCTTAAAGTCACACAAGATGTACACCGGGCTGTGCTTGTGATAA	1201
Qy	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
Db	1202	ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGACTGTGTGGCA	1261
Qy	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle	398
Db	1262	GATGCTTCGCTGAACAACATGAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT	1321
Qy	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAspGlu	418
Db	1322	CCCCAGGGTTCATATGATTCACCTCTTGAAAGGAACGTGTACTTAGTGGGTGATGAA	1381
Qy	419	LysHisArgArgThrTyrAlaAlaArgProThrProAsnAspAspThrLeuAspGluGly	438
Db	1382	AAGCACAGAAAGAACTTACCTCGCGCTCCCACTCCAAATGATGACACTTGGATGAAGA	1441
Qy	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysIleVal	458
Db	1442	GTAGACTTGTGCATTCAAACATAGCAACTGAGCATATTCGAAGCCCTGCCAAGAAATA	1501
Qy	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
Db	1502	CCAAGACTCCCTGCCACAGCAGACAACCTGAAGCAGCTGTCAATGTAATGGGAAACAT	1561
RESULT 5			
ID	ACC62331	standard; cDNA; 1601 BP.	
XX	AC	ACC62331;	
XX	AC	ACC62331;	
DT	23-JUN-2003	(first entry)	
XX	DE	Human NOV41f encoding cDNA SEQ ID NO:191.	
KW	XX	Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;	
KW	KW	anorectic; immunosuppressive; cytostatic; antidiabetic; antinfertility;	
KW	KW	haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;	
KW	KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;	
KW	KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;	
KW	KW	congenital heart defect; aortic stenosis; valve disease; transplanted;	
KW	KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;	
KW	KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;	
KW	KW	fertility; haemophilia; hypercoagulation; graft versus host disease;	
KW	KW	idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;	
KW	KW	Crohn's disease; multiple sclerosis; infectious disease; cancer;	
KW	KW	cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;	
KW	KW	immune disorder; haematopoietic disorder; dyslipidaemia;	
XX	OS	Homo sapiens.	
XX	PN	WO2003023001-A2.	
XX	PD	20-MAR-2003.	
XX	PF	09-SEP-2002; 2002WO-US028538.	

XX 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoili EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Patnirajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zernusen BD, Zhong M;
XX
XX WPI; 2003-313241/30.
DR P-PSDB; ABR54262.
XX
PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
XX
PS Claim 20; Page 242-243; 460pp; English.
XX
XX The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis
CC valve disease, tuberculous sclerosis, scleroderma, obesity, transplantation
CC congenital adrenal hyperplasia, prostate cancer, fertility, haemophilia
CC disorders, neoplasm, lymphoma, uterus cancer, diabetes, metabolic
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1601 BP; 473 A; 321 C; 369 G; 438 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.12e-249
Score: 2480.00
Percent Similarity: 91.92%
Best Local Similarity: 91.92%
Query Match: 98.77%
DB: 7
Length: 1601
Matches: 478
Conservative: 0
MisMatches: 0
Indels: 42
Gaps: 1

US-10-622-516-2 (1-478) x ACC62331 (1-1601)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 13 ATGCCTGGATCATTCTTGAATGCAGAGCTTGCGCCAAAGATGGGAATTGTT 72
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 73 GCCCTGAGATCTATTCTCTCTCAATATGTTGATCAAGCAGAGTTGGAATAATATGAT 132
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 133 GGTTAGATGCTGGAAGTATACCATGGCTTGGCCAGGCCAAGATGGCTTCTGCACA 192
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 193 GATAGAGAAGATATTAACTCTCTTGCAATGACTGTGCTCAGAAATCTTATGAGAGAAAT 252
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 253 AACCTTCTTATGATTGCAATGGCGGGCTGGAAAGTTGGAACAGAGACAATCATCGACAAA 312
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 313 TCAAAGTCTGTGAAGACTTAATTGATGCCAGCTGTTGAAGAGCTCGGGAATACAGATATA 372
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 373 GAAGGATCGACACAACTAAATGCATGCTATGGAAGCAGAGCTGCTGTCTCAATGCTGTT 432
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 433 AACTGATTTGAGTCCAGCTCTGGAGTGAACGGTATGCCCTGGTAGTTGCAGAGATATT 492
QY 149 ----- 149
Db 493 GCTGATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGCAGTAGCTCTGCTA 552
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 553 ATTGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGACACATATGCAACAT 612
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 613 GCCTATGATTTTACAAAGCTGATATGCTAATATCTATAGTAGATGGAATAACTC 672
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 673 TCCATACAGTGTACTACCTCAGTGCATTAGACCGCTGCTACTCTCTACTGCATAAAGATC 732
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 733 CATGCCCAAGTGGCAGAAAGGGAATGATTAAGATTTTACCTTGAATGATTTGGCTTC 792
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 793 ATGATCTTTCACCTCACCATATTGTAACTGTTCAAGAAATCTCTAGCTCGGATGTTGCTG 852
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 853 AATGACTTCCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTGGCCTGGAAGCC 912
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 913 TTTGGGGAATGTTAAATTAGAGACACCTACTTGTAGAGATGTGAGAAAGCATTTATG 972

QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 973 AAGGCTAGCTCTGAACCTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTAACAATCAA 1032
QY 299 AsnGlyAsnMetTyrThrSerSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1033 AATGAAATATGTACACATCTTCAGTATATGTTCCCTTGCAATCTGTTCTAGCACAGTAC 1092
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1093 TCACCTCAGCAATTAGCAGGGAAGAGAAATTGAGTGTCTTATGTTCTGTTGGCT 1152
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1153 GCCACTCTGTACTCTCTTAAAGTCACACAGATGCTACACCGGGTCTGCTCTTGATAAA 1212
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1213 ATAACAGCAAGTTTATGTGATCTTAAATCAAGCTTGATTCAGAAGCTGTTGGCACCA 1272
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db 1273 GATGCTTCGCTGAAAAACAATGAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT 1332
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAspGlu 418
Db 1333 CCCCAGGGTTCAATAGATTCACTCTTTGAAGAAACGTGTGACTTAGTAGGGTGGATGAA 1392
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1393 AAGCACAGAGAAGACTTACCGCTCGGCGTCCCACTCCAATATGACACTTTGGATGAAGGA 1452
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db 1453 GTAGACTTGTGCATTCAACATATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1512
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db 1513 CCAAGACTCCCTGCCACAGACGAGAACCTGAAGCAGCTGTGATTAGTAATGGGGAACAT 1572
RESULT 6
ACC62336 standard; cDNA; 1650 BP.
AC ACC62336;
XX 23-JUN-2003 (first entry)
DT Human NOV41k encoding cDNA SEQ ID NO:201.
DE
XX Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
XX anorectic; immunosuppressive; cytostatic; antidiabetic; antifertility;
XX haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
XX neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;
XX gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
XX congenital heart defect; aortic stenosis; valve disease; transplantation;
XX tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
XX prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
XX fertility; haemophilia; hypercoagulation; graft versus host disease;
XX idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
XX Crohn's disease; multiple sclerosis; infectious disease; cancer;
XX cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
XX immune disorder; haematopoietic disorder; dyslipidaemia;
XX metabolic syndrome X; gene; ss.
OS Homo sapiens.
XX
XX WO2003023001-A2.
PN
XX 20-MAR-2003.
PD
XX 09-SEP-2002; 2002WO-US028538.
PF

XX 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
PA (CURA-) CURAGEN CORP.

XX Agge ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patnrajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zernhusen BD, Zhong M;
XX
XX WPI; 2003-313241/30.
DR P-PSDB; ABR54267.

PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.

XX
PS Claim 20; Page 245; 460pp; English.

XX The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiac,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberos sclerosis, scleroderma, obesity, transplntation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.17e-249 Length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 7 Gaps: 1

US-10-622-516-2 (1-478) x ACC62336 (1-1650)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
DB 22 ATGCCTGCATCACTTCCTTGAATGCAGAAGCTTGCTGGCCAAAGATGTGGGATTGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
DB 82 GCCCTGAGATCTATTTCCTTCTCAATATGTGATCAAGCAGAGTTGAAAAATATGAT 141
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
DB 142 GGTGTAGATGCTGGGAAGTATACCATTTGGCTTGGCCAGCCAAAGATGGCTTGACA 201
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
DB 202 GATAGAGAAGATATTAACTCTCTTGATGATGACTGTGGTTCAAGATCTTATGAGAGAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleLeuAspLys 100
DB 262 AACCTTTCCTATGATTCATTCATTCGCGGCTGGAAAGTTGGACAGACAATCATCGACAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
DB 322 TCAAAGTCGTGAAGACTAATTTGATGCAGCTGTTGAAGTCTGGAAATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
DB 382 GAAGAAATGCACACAACCTAATTCATGATGAGGACACAGCTGCTCTCAATGCTGTT 441
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
DB 442 AACTGATTTGATGTCAGCTCTGGGATGAGCGTATGCCCTGTGATTGCAGAGATATT 501
QY 149 ----- 149
DB 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGACAGTAGCTCTGCTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
DB 562 ATTTGGCCAAATGCTCTTTAATTTTGAACGAGGGCTTGTGGACACATATGCAACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
DB 622 GCTTATGATTTTACCAAGCCTGATATGCTATCTGAATATCTTATAGATGGAAATC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
DB 682 TCCATACAGTGTACCTACCTCAGTCATTAAGACCGCTGCTATCTGTACTGCAAAAAAGATC 741
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
DB 742 CATGCCAGTGGCAGAAAGAGGAAATGATAAGATTTCCTGAATGATTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
DB 802 ATGATCTTTCACCTACCAATATTTGAACTGTTCAGAAATCTCTAGCTCGGATGTTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
DB 862 AATGACTTCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTGCGCTGAAGCC 921
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
DB 922 TTTGGGATGTTAAATTAGAGACACCTACTTTGATAGATGTGTGAGAGGACATTTATG 981

QY	279	LYSALA	SerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSer	AngIn	298				
Db	982	AAGCTAGCTCTGA	CTCTTCACTCAGTACAGAAAACAAGCATCTTTACTTGATCAATCA		1041				
QY	299	ASNGIYA	SNMeTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr		318				
Db	1042	AATGAAATATGTAC	ACACTTCTCAGTATATGGTTCCCTTGACATCTGTCTAGCACAGTAC		1101				
QY	319	SerProGlnGlnLeu	AlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla		338				
Db	1102	TCACCTCAGCAAT	TAGCAGGAAAGAAATGGAGTGTTCCTTATGGTTGGCT		1161				
QY	339	AlaThrLeuTyrSer	LeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys		358				
Db	1162	GCCACTCTGTACT	CTCTTTAAAGTCACACACAGATGCTACACCGGGCTCTCTTGATATA		1221				
QY	359	IleThrAlaSerLeu	CysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro		378				
Db	1222	ATAACAGCAAGTT	ATATGTGATCTTAAATCAAGGCTTGATTCAAGACTGGTGGCACCA		1281				
QY	379	AspValPheAlaGlu	AsnMeLysLeuArgGluAspThrHisIleLeuValAsnTyrIle		398				
Db	1282	GATGCTCTCGCT	GAAAAACATGAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT		1341				
QY	399	ProGlnGlySerIle	AspSerLeuPheGluGlyThrTryptyrLeuValArgValAspGlu		418				
Db	1342	CCCCAGGGTTCAT	AGATGATCACTCTTGAAGACAAGTGTACTTATGTTAGGGTGATGAA		1401				
QY	419	LysHisArgArgThr	TyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly		438				
Db	1402	AAGCACAGAAACT	TACGCTCGCGCTCCCACTCCAAATGATGACACTTTGGATGAAGCA		1461				
QY	439	ValGlyLeuValHis	SerAsnIleAlaThrGlnHisIleProSerProAlaLysLysVal		458				
Db	1462	GTAGGACTTGTGC	ATTCAACATAAGCACTGAGCATATTCCAGCCCTGCCAAGAAAGTA		1521				
QY	459	ProArgLeuProAla	ThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis		478				
Db	1522	CCAAGACTCCCTG	CCACACAGACGAACCTGAAGCAGCTGTCATTAGTAATGGGAACAT		1581				
RESULT 7									
ID ACC62328 standard; cDNA; 1650 BP.									
XX	AC	ACC62328;							
XX	DT	23-JUN-2003 (first entry)							
DE	Human NOV41c encoding cDNA SEQ ID NO:185.								
XX	Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;								
KW	anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;								
KW	haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;								
KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;								
KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;								
KW	congenital heart defect; aortic stenosis; valve disease; transplantation;								
KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;								
KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;								
KW	fertility; haemophilia; hypercoagulation; graft versus host disease;								
KW	idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;								
KW	Crohn's disease; multiple sclerosis; infectious disease; cancer;								
KW	cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;								
KW	immune disorder; haematopoietic disorder; dyslipidaemia;								
KW	metabolic syndrome X; gene; ss.								
XX	Homo sapiens.								
OS	Homo sapiens.								
XX	PN	WO2003023001-A2.							
XX	PD	20-MAR-2003.							
PF	09-SEP-2002; 2002WO-US028538.								

07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.

(CURA-) CURAGEN CORP.

Aggee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gussev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patrajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zerhusen BD, Zhong M;

WPI; 2003-313241/30.
DR P-PSDB; ABR54259.

Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.

Claim 20; Page 241; 460pp; English.

The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaemic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis
CC valve disease, tuberosus sclerosis, scleroderma, obesity, diabetes, metabolic
CC congenital adrenal hyperplasia, prostate cancer, fertility, haemophilia,
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, cancer-associated cachexia, cancer,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
XX the human NOV35b protein in the exemplification of the present invention

Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.17e-249	Length:	1650
Score:	2480.00	Matches:	478
Percent Similarity:	91.92%	Conservative:	0
Best Local Similarity:	91.92%	Mismatches:	0
Query Match:	98.77%	Indels:	42
DB:	7	Gaps:	1

US-10-622-516-2 (1-478) x ACC62328 (1-1650)

QY	1	MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLyAspValGlyIleVal	20
DB	22	ATGCCTGGATCACTTCCTTGAATGCAGAGCTTGCTGGCCAAAAGATGTTGGGATTGTT	81
QY	21	AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp	40
DB	82	GCCCTTGAGATCTATTTCTCTCAATATGTTGATCAAGCAGAGTTGAAAAATATGAT	141
QY	41	GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
DB	142	GGTGTAGATGCTGGAGATATACATTGGCTTGGCCAGGCCAAGATGGCTTCTGCACA	201
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn	80
DB	202	GATAGAGAAGATATTAACTCTCTTGCATGACTGTGTTCAGAATCTTATGAGAGAAAT	261
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
DB	262	AACCTTTCCTATGATTGCATTGGCGCGCTGGAAGTTGGAACAGACAAATCATCGACAAA	321
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
DB	322	TCAAACTCTGTGAAGACTAATTGATGCAGCTGTTGAAGACTCTGGAAATACAGATATA	381
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaIleValPheAsnAlaVal	140
DB	382	GAAAGAAATGCACAACTAATGCTATGAGGACACAGCTGCTCTTCAATGCTGTT	441
QY	141	AsnTrpIleGluSerSerSerTrpAsp-----	149
DB	442	AACTGATTGATCCAGCTCTGGGATGGAACGATATGCCCTGGTAGTTGCAGAGATATT	501
QY	149	-----	149
DB	502	GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGCAGTAGCTTGCTA	561
QY	150	-----GlyLeuArgGlyThrHisMetGlnHis	158
DB	562	ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT	621
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu	178
DB	622	GCTATGATTTTACAAAGCCTGATATGCTATCTGAATATCTTATAGTAGATGGAAATCTC	681
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
DB	682	TCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTATTCTCTACTGCAAAAAGATC	741
QY	199	HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
DB	742	CATGCCCACTGGCAGAAAGAGGAAATGATAAAGATTTTACCTTGAATGATTTGGCTTC	801
QY	219	MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu	238
DB	802	ATGATCTTTCACTCACCATATTTGTAACGTGTTCAGAAATCTTAAGCTCGGATGTTGCTG	861
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla	258
DB	862	AATGACTTCTTAATGACCAGAAATAGAGATAAAAATAGTATCTATAGTGGCTTGAAGCC	921
QY	259	PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet	278
DB	922	TTTGGGATGTTAAATTAGAAAGACCTACTTGTATAGAGTGTGGAAGAGCATTTTATG	981

QY	279	LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln	298
DB	982	AAGCTAGCTCTGAACCTTTCAGTCAGAAAACAAAGGCATCTTACTGTATCAAAATCAA	1041
QY	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
DB	1042	AATGAAATATATGACACATCTTCAGTATATGTTCCCTTGCAATCTGTTCTAGCACAGTAC	1101
QY	319	SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
DB	1102	TCACTCAGCAATTAGCAGGGAAGAAATTGAGTGTTTTCTTATGTTCTGTGTTGGCT	1161
QY	339	AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys	358
DB	1162	GCCACTCTGTACTCTCTTAAAGTCACACAGATGCTACACCGGGTCTGCTTGATATAA	1221
QY	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
DB	1222	ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTGTGTGGCACCA	1281
QY	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisLysLeuValAsnTyrIle	398
DB	1282	GATGCTTCGCTGAAAAACATGAAAGCTCAGAGAGGACCCATCATTTGTCAACTATATT	1341
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu	418
DB	1342	CCCCAGGTTCAATAGATCACTCTTGAAGGAACGTGTACTTAGTGGTGATGAA	1401
QY	419	LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly	438
DB	1402	AAGCAGAGAAGAACTTACGCTCGGGCTCCCACTCCAATGATGACACTTGGATGAAGGA	1461
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal	458
DB	1462	GTAAGACTTGTGCATTCAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAGTA	1521
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
DB	1522	CCAAAGACTCCCTGCCACAGCAGCAAACTGAAGCAGCTGTCAATTAGTATGGGGAACAT	1581
RESULT 8			
ID	ACC62326	standard; cdna; 1650 BP.	
AC	ACC62326;		
DT	23-JUN-2003	(first entry)	
XX	Human NOVA1a encoding cDNA SEQ ID NO:181.		
DE			
XX			
KW	Human; NOVA; antiatherosclerotic; hypotensive; cardiant; dermatological;		
KW	anorectic; immunosuppressive; cytostatic; antidiabetic; antifertility;		
KW	haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;		
KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;		
KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;		
KW	congenital heart defect; aortic stenosis; valve disease; transplantation;		
KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;		
KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;		
KW	fertility; haemophilia; hypercoagulation; graft versus host disease;		
KW	idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;		
KW	Crohn's disease; multiple sclerosis; infectious disease; cancer;		
KW	cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;		
KW	immune disorder; haematopoietic disorder; dyslipidaemia;		
KW	metabolic syndrome X; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO2003023001-A2.		
XX			
PD	20-MAR-2003.		
XX			
PF	09-SEP-2002; 2002WO-US028538.		

XX 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
PA (CURA-) CURAGEN CORP.

XX Agge ML, Alsobrook JP, Anderson DW, Berghs C, Boldog EV,
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A,
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K,
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W,
PI Kerkda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X,
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Patturajan M,
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA,
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ,
PI Zernusen BD, Zhong M;
XX
DR WPI; 2003-313241/30.
DR P-PSDB; ABR54257.

XX Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.

XX Claim 20; Page 240; 460pp; English.

XX The present invention describes isolated human NOX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOX proteins given in
CC ABR54167 to ABR54276. NOX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC antiaesthetic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipemic activities, and can be used in gene
CC therapy. NOX proteins are useful for treating or preventing a pathology
CC associated with a NOX protein in humans and for treating a syndrome
CC associated with the human disease. NOX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valvular disease, tubercular scleriosis, scleroderma, obesity, diabetes, metabolic
CC congenital adrenal hyperplasia, prostate cancer, diabetes, transplantion,
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.17e-249 Length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 7 Gaps: 1

US-10-622-516-2 (1-478) x ACC62326 (1-1650)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProlYsAspValGlyIleVal 20
Db 22 ATGCCTGGATCACCCTCTTGAATGCAGAACTTGCTGGCCAAAAGATGTGGCATTT 81
QY 21 AlaleuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluIlyTyrAsp 40
Db 82 GCCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAATATGAT 141
QY 41 GlyValAspAlaGlyIlyTyrThrIleGlyLeuGlyGlnAlaIlyMetGlyPheCysThr 60
Db 142 GGTGTAGATCTGGGAGATATACCATTTGGCTTGGGCCAGCCAGATGGGCTTGCACA 201
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATTAACTCTTCTTGATGACTGTGGTTCAGAATCTTATGAGAGAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspIly 100
Db 262 AACCTTTCCTATGATTCGATTCGGCGGCTGGAAGTTGGAACAGACAATCATGCACAA 321
QY 101 SerIlySerValIlyThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAAGTGTGGAAGACTTAATTTGATGACGCTGTTGAAGGTCTGGAAATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 382 GAAGGAATGCACACAACCTAATGCATGTGAGGACACAGCTGCTGTCTCAATGCTGTT 441
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 442 AACTGATGATGATCCAGCTCTTGGGATGACGGTATGCCCTGTAGTTGCAGAGATATT 501
QY 149 ----- 149
Db 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGACAGTAGCTGCTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT 621
QY 159 AlaTyrAspPheTyrIlyAspProAspMetLeuSerGluTyrProIleValAspGlyIlyLeu 178
Db 622 GCCTATGATTTTACAAAGCCTGATATGCTATCTGAATATCTATAGTAGATGGAAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysIlyIlySile 198
Db 682 TCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTACTGCAAAAAGATC 741
QY 199 HisAlaGlnTrpGlnIlySgIlyAsnAspIlyAspPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCAGTGCAGAAAGAGGAAATGATTAAGATTTAACCTTGAATGATTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysIlySleuValGlnIlySerIleuAlaArgMetLeuLeu 238
Db 802 ATGATCTTCACTCACCAATATTTAACTGGTTCAGAATCTCTAGCTCGGATGTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspIlyAsnSerIleTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCTTAATGACCAAGATAGAGATMAAAATAGTATCTATAGTGGCCTGGAAGCC 921
QY 259 PheGlyAspValIlySleuGluAspThrTyrPheAspArgAspValGluIlyAlaPheMet 278
Db 922 TTGGGGATGTTAAATTAGAGACACCTACTTGTATAGATGTGGAAGAGCATTTATG 981

QY	279	LySAlaSerSerGlnLeuPheSerGlnLyThrLySAlaSerLeuLeuValSerAsnGln	298
Db	982	AAGGCTAGCTCTGAAGCTTCTTCAAGTCAAGAAACAAAGGCATCTTTACTTGATCAATCA	1041
QY	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
Db	1042	AATGAAATATATGTACACATCTTCAGTATATAGTTCCCTTGATCTGTTCTAGACAGTAC	1101
QY	319	SerProGlnGlnLeuAlaGlyLySArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
Db	1102	TCACCTCAGCAATTAGCAGGGAAGAAGAAATTGAGTGTCTTTCTTATAGTTCTGATTGGCT	1161
QY	339	AlaThrLeuTyrSerSerLeuLySValThrGlnAspAlaThrProGlySerAlaLeuAspLyS	358
Db	1162	GCCACTCTGTACTCTCTTAAGTCAACACAGATGCTACCCGGGCTGCTCTTGATATA	1221
QY	359	IleThrAlaSerLeuCysAspLeuLySAspSerArgLeuAspSerArgThrGlyValAlaPro	378
Db	1222	ATAACAGCAAGTTTATGTGATCTTAATCAAGGCTTGATTCAGAAGACTGTTGGCACCA	1281
QY	379	AspValPheAlaGluAsnMetLySLeuArgGluAspThrHisIleLeuValAsnTyrIle	398
Db	1282	GATGCTTCGCTGAACAATGAGCTCAAGCTCAGAGAGACACCATCATTTGGTCAACTATATT	1341
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu	418
Db	1342	CCCCAGGTTCAATGATTCATCTTTGAAGGAACGTGTACTTAGTGGTGATGAA	1401
QY	419	LySHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluIy	438
Db	1402	AAGCACAGAAAGAACTTACGCTCGCGCTCCACTCCAATGATGACACTTTGATGAAGA	1461
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLySLeuVal	458
Db	1462	GTAGGACTTGTGCATTCAAACATAGCACTGAGCATATTCCAAGCCCTGCCAAGAAGTA	1521
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
Db	1522	CCAAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAGCTGTCTATTAGTATGGGAACAT	1581
RESULT 9			
ID	ACC62330	standard; cDNA; 1650 BP.	
XX	AC	ACC62330;	
XX	DT	23-JUN-2003 (first entry)	
DE	XX	Human NOV41e encoding cDNA SEQ ID NO:189.	
XX	XX	Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;	
KW	KW	anorectic; immunosuppressive; cyostatic; antidiabetic; antinfertility;	
KW	KW	haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;	
KW	KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;	
KW	KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;	
KW	KW	congenital heart defect; aortic stenosis; valve disease; transplantation;	
KW	KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;	
KW	KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;	
KW	KW	fertility; haemophilia; hypercoagulation; graft versus host disease;	
KW	KW	idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;	
KW	KW	Crohn's disease; multiple sclerosis; infectious disease; cancer;	
KW	KW	cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;	
KW	KW	immune disorder; haematopoietic disorder; dyslipidaemia;	
XX	XX	metabolic syndrome X; gene; ss.	
OS	XX	Homo sapiens.	
XX	XX	WO2003023001-A2.	
PN	XX	20-MAR-2003.	
XX	XX		
PD	XX	09-SEP-2002; 2002WO-US028538.	
PF			

XX	07-SEP-2001;	2001US-0318120P.
PR	07-SEP-2001;	2001US-0318184P.
PR	10-SEP-2001;	2001US-0318430P.
PR	17-SEP-2001;	2001US-0322636P.
PR	17-SEP-2001;	2001US-0322781P.
PR	17-SEP-2001;	2001US-0322816P.
PR	17-SEP-2001;	2001US-0322817P.
PR	19-SEP-2001;	2001US-0323519P.
PR	20-SEP-2001;	2001US-0323631P.
PR	20-SEP-2001;	2001US-0323636P.
PR	25-SEP-2001;	2001US-0324969P.
PR	25-SEP-2001;	2001US-0325091P.
PR	26-SEP-2001;	2001US-0324990P.
PR	14-DEC-2001;	2001US-0341144P.
PR	26-FEB-2002;	2002US-0351959P.
PR	05-MAR-2002;	2002US-0361663P.
PR	03-MAY-2002;	2002US-0377908P.
PR	17-MAY-2002;	2002US-0381483P.
PR	29-MAY-2002;	2002US-0383863P.
PR	02-JUL-2002;	2002US-0393332P.
PR	17-JUL-2002;	2002US-0396412P.
PR	13-AUG-2002;	2002US-0403517P.
PR	06-SEP-2002;	2002US-00236417.

(CURA-) CURAGEN CORP.

PI Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Patturajan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zernusen BD, Zhong M;

WPI; 2003-313241/30.
P-PSDB; ABR54261.

Novel human proteins and nucleic acid encoding the proteins, useful for diagnosis, treatment and prevention of disorders involving the human protein or nucleic acid e.g. cardiac and neurological disorders.

Claim 20; Page 242; 460pp; English.

The present invention describes isolated human NOVX proteins, where X is 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiatic, hypotensive, dermatological, anorectic, immunosuppressive, cytostatic, antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV, antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic, antiparkinsonian and antilipaeamic activities, and can be used in gene therapy. NOVX proteins are useful for treating or preventing a pathology associated with a NOVX protein in humans and for treating a syndrome associated with the human disease. NOVX nucleic acids, proteins and antibodies can be used in the treatment and diagnosis of cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, valve disease, tuberosus sclerosis, scleroderma, obesity, transplantation, congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis, infectious disease, anorexia, cancer-associated cachexia, cancer, Alzheimer's disease, Parkinson's disease, immune disorders, haematopoietic disorders, dyslipidaemias, and metabolic syndrome X. ACC62346 to ACC62465 represent PCR primers and probes for human NOVX sequences, which are used in examples from the present invention. ABR54277 represents a human trypsinogen protein given in comparison with the human NOV35b protein in the exemplification of the present invention

Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Alignment Scores: Pred. No.: 1.17e-249 Length: 1650 Score: 2480.00 Matches: 478 Percent Similarity: 91.92% Conservative: 0 Best Local Similarity: 91.92% Mismatches: 0 Query Match: 98.77% Indels: 42 DB: 7 Gaps: 1

US-10-622-516-2 (1-478) x ACC62330 (1-1650)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLySAspValGlyIleVal 20
Db 22 ATGCCTGCATCACTTCCTTGAATGCAGAGCTTGCTGGCCAAAAGATGTTGGATTGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 82 GCCCTTGAGATCTATTTCTCTCTCAATATGTTGATCAAGCAGAGTGGAAAAATATGAT 141
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 142 GGTGTAGATGCTGGGAAGTATACCATTTGGCTGGCCAGGCCAAGATGGGCTTCTGCACA 201
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATTAACTCTTGGCATGACTGTGTTCAAGATCTTATGAGAGAAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 262 AACCTTCTATGATTGCAATGGCGCGGTGGAAGTGGAAACAGACAAATCATCGACAAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAACTGTGTGAAGACTAAATTGATGCAGCTGTTGAAGAGTCTGGGAATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 382 GAAGGAATGCACACAACATAATGCTATGAGGACACAGCTGCTGTTCAATGCTGTT 441
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 442 AACTGATGAGTCCAGCTTTGGGATGACGCTATGCCCTGTAGTTCAGAGAGATATT 501
QY 149 ----- 149
Db 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTGGAGCAGTAGCTTGCTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 622 GCCTATGATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAAAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 682 TCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTACTGCAAAAAAGATC 741
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCCAAGTGGCAGAAAGAGGAAATGATTAAGATTTTACCTTGAATGATTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 802 ATGATCTTTCACTCACCATATTTGTAACCTGTTCAGAATCTTAGCTCGGATGTTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCCTTAATGACCAAGATAGAGATAAAATAATGATATAGTGCGCTGGAAGCC 921
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 922 TTTGGGAGATGTTAAATTAGAAAGCACCTACTTGTATGAGAGATGTGAGAAAGCATTATG 981

QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 982 AAGGCTAGCTCTGAACCTTTCAGTACAGAAACAAGGCATCTTACTTGATCAATCAA 1041
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1042 AATGAAATATGTACACATCTTCAGTATATGTTCCCTTGCAATCTGTCTAGCACAGTAC 1101
QY 319 SerProGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1102 TCACCTCAGCAATTAGCAGGGAAGAAATGGAGTGTTCCTTATGTTCTGTGGCT 1161
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1162 GCCACTCTGTACTCTCTTAAAGTCACACAGATGCTACACCGGGGTCTGCTCTTGATATAA 1221
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1222 ATAACAGCAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAGAACTGTGTGSCACCA 1281
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle 398
Db 1282 GATGCTTCGCTGAAAAACATGAAGCTCAGAGAGACACCCATCATTTGTCATATATT 1341
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAspGlu 418
Db 1342 CCCAAGGTTCAATAGATTCACTCTTGAGAGAAAGTGTACTTAGTGGTGCGATGAA 1401
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1402 AAGCACAGAAAGAACTTACGCTCGGCGTCCCACTCCAATGATGACACTTGGATGAAGGA 1461
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db 1462 GTAGACTTGTCATTCAACATAGCAACTGAGCATATTCCAAAGCCCTGCCAAAGAAATA 1521
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyLuhis 478
Db 1522 CCAAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAAGCTGTCTATTAGTAATGGGAAACAT 1581
RESULT 10
ACC62338
ID ACC62338 standard; cDNA; 1650 BP.
XX
AC ACC62338;
XX
DT 23-JUN-2003 (first entry)
XX
DE Human NOVA1m encoding cDNA SEQ ID NO:205.
XX
KW Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cytostatic; antidiabetic; antifertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasia; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X; gene; ss.
OS Homo sapiens.
XX
XX
PN WO2003023001-A2.
XX
XX 20-MAR-2003.
PD
XX 09-SEP-2002; 2002WO-US028538.
PF

XX 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.

XX (CURA-) CURAGEN CORP.

XX Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W,
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyanakar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patnurajan M,
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zerhusen BD, Zhong M;
XX WPI; 2003-313241/30.
DR P-PSDB; ABR54269.

XX Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.

XX Claim 20; Page 246; 460pp; English.

XX The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiac,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antifertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaeamic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberos sclerosis, scleroderma, obesity, transplacental,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human tryptanogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.17e-249 Length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 7 Gaps: 1

US-10-622-516-2 (1-478) x ACC62338 (1-1650)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 22 ATGCCTGGATCACTTCCTTGAATGCAGAAGCTTGCTGCCAAAGATGTGGATTGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 82 GCCCTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCAGAGTTGAAAAATATGAT 141
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 142 GGTGTAGATGCTGGAGATATACCATTTGGCTTGGCCAGGCCAAGATGGCTTTCGACA 201
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATTAACTCTCTTGCACTGACTGTGTTCAAGATCTTATGAGAGAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 262 AACCTTCTTATGATTGCTATTTGGCGGCTGGAAGTTGGAACAGACAATCATCGACAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAGTCTGTGAAGACTTAATTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 382 GAAGAAATCGACACAACATAATGCATGCTATGAGAGCAGACTGCTCTTCAATGCTGTT 441
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 442 AACTGATGTAGTCCAGCTCTTGGATGAGCGTATGCCCTGTAGTTGCAGAGATATT 501
QY 149 ----- 149
Db 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGATTGGAGCAGTAGCTTCTGTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATTGGGCCAAATGCTCCTTAAATTTTGAACGAGGCTTCGTGGACACATATATGACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 622 GCTATGATTTTACAAAGCTGATATGCTATCTGAATATCTTATAGTAGAGGAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIle 198
Db 682 TCCATACAGTGTACCTACCTACGTGATAGACCGCTGCTATTCTGTACTGCAAAAAGATC 741
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCCAGTGGCAGAAAGAGGAAATGATTAAGATTTAACCTTGAATGATTTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 802 ATGATCTTCACTACCATATTTGTAACCTGGTTCAAAATCTCTAGCTCGGATGTTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCTTAATGACAGAAATAGATAAAAATATATCTATAGTGCGCTGGAAGCC 921
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGlyLysAlaPheMet 278
Db 922 TTTGGGATGTTAAATTGAAGACACCTACTTGTATGAGATGTGAGAGGCAATTATG 981

QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
| | | | |
Db 982 AAGGCTAGCTCTGAACCTCTTCAGTCAGAAAAACAAGGCATCTTTACTTGTATCAATCAA 1041
| | | | |
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
| | | | |
Db 1042 AATGGAATATATGACACATCTTCAGTATATGTTCCCTTGCACTGTCTTAGCAGACTAC 1101
| | | | |
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
| | | | |
Db 1102 TCACCTCAGCAATTAGCAGGAGGAAGAAATTGAGTGTCTTCTTATGTCTTGATTGGCT 1161
| | | | |
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
| | | | |
Db 1162 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACGGGGTCTGCTCTTGATAAA 1221
| | | | |
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
| | | | |
Db 1222 ATAACAGCAGATTATGTGATCTTAAATCAAAGGCTTGATCAAGAACTGTGTGGCACA 1281
| | | | |
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
| | | | |
Db 1282 GATGCTTCGCTGCAAAACATGAAGCTCAGAGAGACACCCCATCATTTGCTCACTATATT 1341
| | | | |
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
| | | | |
Db 1342 CCCCAGGGGTTCAATGATTCCTACTCTTTGAAGGAACGTGTACTTAGTTAGGTTGATGAA 1401
| | | | |
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
| | | | |
Db 1402 AAGCACAGAGAAGAACTTAACGCTGGCGTCCCACTCCAATGATGACACTTGGATGAAGGA 1461
| | | | |
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
| | | | |
Db 1462 GTAGGACTTGTGCATCAAAACATAGCAACTGAGCATATTCCAAAGCCCTGCCAAAGAAAGTA 1521
| | | | |
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
| | | | |
Db 1522 CCAAGACTCCCTGCCACAGCAGACAGAACTGAAGCAGCTGTCTATTAGTAATGGGGAACAT 1581
| | | | |
RESULT 11
ACC62334
ID ACC62334 standard; cDNA; 1650 BP.
XX
AC ACC62334;
XX
DT 23-JUN-2003 (first entry)
XX
DE Human NOV411 encoding cDNA SEQ ID NO:197.
XX
XX Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tubercous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostrate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X; gene; ss.
XX
XX Homo sapiens.
XX
PN WO2003023001-A2.
XX
XX 20-MAR-2003.
PD
XX 09-SEP-2002; 2002WO-US028538.
PF

XX 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
PA (CURA-) CURAGEN CORP.
XX
XX Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton B, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K; Ji W;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Paturajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RU, Twomlow N, Vernet CAM, Voss EZ;
PI Zerhusen BD, Zhong M;
XX
DR WPI; 2003-313241/30.
DR P-PSDB; ABR54265.
XX
XX Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS Claim 20; Page 244; 460pp; English.
XX
XX The present invention describes isolated human NOX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOX proteins given in
CC ABR54167 to ABR54276. NOX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaemic activities, and can be used in gene
CC therapy. NOX proteins are useful for treating or preventing a pathology
CC associated with a NOX protein in humans and for treating a syndrome
CC associated with the human disease. NOX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tubercous sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.17e-249	Length:	1650
Score:	2480.00	Matches:	478
Percent Similarity:	91.92%	Conservative:	0
Best Local Similarity:	91.92%	Mismatches:	0
Query Match:	98.77%	Indels:	42
DB:	7	Gaps:	1

US-10-622-516-2 (1-478) x ACC62334 (1-1650)

QY	1	MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal	20
DB	22	ATGCCCTGCATCACTTCCTTGAATGCAGAGCTTGCTGGCCAAAGATGTTGGGATGTT	81
QY	21	AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp	40
DB	82	GCCCTTGAGATCTATTTCTTCTCAATATGTGATCAAGCAGAGTTGAAAAATATGAT	141
QY	41	GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
DB	142	GGTGTAGATGCTGGGAAGTATACATTGGCTTGGCCAGGCCAAGATGGGCTTGCACA	201
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn	80
DB	202	GATAGAGAAGATATTAACTCTCTTGATGACTGTGTTCAGAATCTTATGAGAGAAAT	261
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
DB	262	AACCTTTCCTATGATGTGCATGGCGCGCTGGAAGTTGGAACAGACAAATCATGCACAA	321
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
DB	322	TCAAACTCTGTGAAGACTAATTGTGATGCAGCTGTTGAAGAGTCTGGGAATACAGATATA	381
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
DB	382	GAAAGAAATGCACAACTAATGCTATGAGGACACAGCTGCTTCAATGCTGTT	441
QY	141	AsnTrpIleGluSerSerSerTrpAsp-----	149
DB	442	AACTGGATTGATCCAGCTCTGGGATGGAACGCTATGCCCTGTGATTGACAGAGATATT	501
QY	149	-----	149
DB	502	GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGACAGCAGTAGCTTGCTA	561
QY	150	-----GlyLeuArgGlyThrHisMetGlnHis	158
DB	562	ATTGGGCCCAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT	621
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu	178
DB	622	GCCTATGATTTTACAAAGCCTGATATGCTATCTGAATATCTTATAGTAGATGGAAATCTC	681
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
DB	682	TCCATACAGTGTACTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTACTGCAAAAAGATC	741
QY	199	HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
DB	742	CATGCCACAGTGGCAGAAAGAGGAAATGATAAAGATTTTACCCTTGAATGATTTGGCTTC	801
QY	219	MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu	238
DB	802	ATGATCTTTCACTCACCATATTTGTAACGTGTTCAGAAATCTTACGCTCGGATGTTGCTG	861
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla	258
DB	862	AATGACTTCTTAATGACCAGAAATAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC	921
QY	259	PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet	278
DB	922	TTTGGGATGTTAAATTAGAAAGACCTACTTGTATAGAGATGTGGAAGGCAATTTATG	981

QY	279	LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln	298
DB	982	AAGGCTAGCTTGAACTCTTCAGTCAGAAACAAAGGCATCTTACTGTATCAAAATCAA	1041
QY	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
DB	1042	AATGAAATATATGACACATCTTCAGTATATAGTTCCCTTGCAATCTGTTCTAGACAGTAC	1101
QY	319	SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
DB	1102	TCACTCAGCAATTAGCAGGGAAGAAATTGAGTGTGTTTCTTATGTTCTGTGTTGGCT	1161
QY	339	AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys	358
DB	1162	GCCACTCTGTACTCTTAAAGTCAACAGATGCTACACGGGGTCTGCTTGATAA	1221
QY	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
DB	1222	ATAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTGTGGACCA	1281
QY	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle	398
DB	1282	GATGCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCACATCTATTGTCAACTATATT	1341
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu	418
DB	1342	CCCCAGGTTCAATAGATCACTCTTGAAGGAACGTGTACTTAGTGGTGATGAA	1401
QY	419	LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly	438
DB	1402	AAGCAGAGAAGAACTTAGCGTCGGCGTCCCACTCCAATGATGACACTTTGATGAAGGA	1461
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal	458
DB	1462	GTAGGACTTGTGCATTCAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAGTA	1521
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
DB	1522	CCAAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAGCTGTCAATTAGTAATGGGGAAT	1581

RESULT 12
ACC62327
ID ACC62327 standard; cDNA; 1650 BP.
AC ACC62327;
XX 23-JUN-2003 (first entry)
DT
XX
DE Human NOV41b encoding cDNA SEQ ID NO:183.
XX
KW Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cytostatic; antidiabetic; antinfertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuoprotective; nootropic; antiparkinsonian; metabolic; antilipaeamic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003023001-A2.
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028538.

XX 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX
PA (CURA-) CURAGEN CORP.

XX Agee ML, Alsebrook JP, Anderson DW, Berghs C, Boldog FL,
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dhippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kerkuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malvankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI Pena CE, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zernusen BD, Zhong M;

DR WPI; 2003-313241/30.
DR P-PSDB; ABR54258.

XX Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.

XX Claim 20; Page 240-241; 460pp; English.

XX The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antiinfectivity, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipemic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tubercous sclerositis, scleroderma, obesity, transplacental,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human tyrosinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.17e-249 Length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 7 Gaps: 1

US-10-622-516-2 (1-478) x ACC62327 (1-1650)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 22 ATGCTGGATCATTCTTCTTGAATGCAGAGCTTGCTGGCCAAAGATGTTGGATTGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 82 GCCCTTGAGATCTATTTCCTTCCTCAATATGTTGATCAAGCAGAGTGGAAAAATATGAT 141
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 142 GGTGTAGATGCTGGGAAGTATACCATTTGGCTTGGCCAGGCCAAGATGGCTTTCACACA 201
QY 61 AspArgGluAspIleAsnSerLeuCyMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATATCTCTCTTGGCATGACTGTGGTTCAAGATCTTATGAGAGAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyTyrGluThrIleIleAspLys 100
Db 262 AACCTTTCCTATGATTGCATTGGCGCGGTGGAAGTTGGAACAGAGACAATCATCGACAAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAAAGTCTGGAAGACTTAATTGATGACGCTGTTTGAAAGAGCTCGGGAATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyTyrAlaAlaValPheAsnAlaVal 140
Db 382 GAAGGAATGCACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCTTCAATGCTGTT 441
QY 141 AsnTrpIleGluSerSerTyrAsp----- 149
Db 442 AACTGGATTGAGTCCAGCTCTTGGATGGACGGTATGCCCTGTAGTTGCAGAGATATT 501
QY 149 ----- 149
Db 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGAGCAGTAGCTCTGCTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATTGGCCAAATGCTCTTTAATTTTGAACGAGGCTTCGTGGGACACATATGCAACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 622 GCCTATGATTTTACAAAGCCTGATATGCTAATATCTATAGTAGATGGGAAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 682 TCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTACTGCAAAAAGATC 741
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCAGTGCAGAAAAGGAGAAATGATTAAGATTTCCTTGAATGATTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 802 ATGATCTTCTACTCACCATATATGTAACCTGGTTCAGAAATCTCTAGCTCGATGTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCTTAATGACCAGAAATAGAGATAAATAGTATCTATAGTGGCTGGAAGCC 921
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 922 TTGGGGATGTTAATTAGAAACACCTACTTGTATAGAGATGTGGAAGGCAATTATATG 981

QY	279	LYSALASERSEGLULEUPHESERGLNLYTHRLYSALASERLEULEVALSERASGLN	298
DB	982	AAGGCTAGCTTGAACCTTTCAGTCAGAAAACAAGGCATCTTACTGTATCAATCAA	1041
QY	299	ASNGLYASNMEITYRTHRSESERVALTYRGLYSEXTLEUALSERVALLEUALAGINTYR	318
DB	1042	AATGAAATATGTACACATCTTCAGTATATGTTCCCTGCATCTGTTCTAGACACAGTAC	1101
QY	319	SERPROGLNGLNLEUALGLYLYSARGILEGLYVALPHESERTYRGLYSERGLYLEUALA	338
DB	1102	TCACCTCAGCAATTAGCAGGAGAAGAAATTGGAGTGTTTCTTATGTTCTGTGTTGGCT	1161
QY	339	ALATHRLEUTYRSEIRLEULYSVALTHRGLNASPALATHRPROGLYSERALEUALSPLYS	358
DB	1162	GCCACTCTGTACTCTTTAAAGTCACACACAGATGCTACACGGGGCTCTGCTCTTGATAAA	1221
QY	359	ILETHRALASERLEUCYASBPLEULYSSEARGLLEUASPSERARGTHRGLYVALALAPRO	378
DB	1222	ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATCAAGAACTGTTGGCACCA	1281
QY	379	ASPVALLPHEALAGLUASNMEITYSLEUARGLUASPTHRHISHISLEUVALASNTRYILE	398
DB	1282	GATGCTTCGCTGAANAACATGAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT	1341
QY	399	PROGLNGLYSERILEAPSERLEUPHEGLUGLYTHRTPFYLEUVALARGVALASPGLU	418
DB	1342	CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGTACTTAGTTAGGGTGATGAA	1401
QY	419	LYSHISARGARGTHRTRYALAAARGARGPROTHRPROASNAPSPTHRLEUASPGUGLY	438
DB	1402	AAGCACAGAAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTTTGATGAAGGA	1461
QY	439	VALGLYLEUVALHISSEASNNLEALATHRGLNHISILEPROSERPROALALYSLYVAL	458
DB	1462	GTAGGACTTGTGCATTCAACATAGCAACTGAGCATATTCCAAAGCCCTGCCAAGAAAGTA	1521
QY	459	PROARGLEUPROALATHRALAALAGLUPROGLUALALAVALLSESASNGLYGLNHIS	478
DB	1522	CCAAAGACTCCCTGCCACAGCAGCAAGAACTGAAAGCAGCTGTGATTAAGTAATGGGAACAT	1581

RESULT	ID	ACC62332	standard; cDNA; 1650 BP.
XX	AC	ACC62332;	
XX	DT	23-JUN-2003	(first entry)
XX	DE	Human NOV41g	encoding cDNA SEQ ID NO:193.
KW		Human; NOX; antiatherosclerotic; hypotensive; cardiant; dermatological;	
KW		aneurotic; immunosuppressive; cyostatic; antidiabetic; antiinfertility;	
KW		haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;	
KW		neuroprotective; nootropic; antiparkinsonian; metabolic; antilipemic;	
KW		gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;	
KW		congenital heart defect; aortic stenosis; valve disease; transplantation;	
KW		tuberosus sclerosis; obesity; congenital adrenal hyperplasia; diabetes;	
KW		prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;	
KW		fertility; haemophilia; hypercoagulation; graft versus host disease;	
KW		idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;	
KW		Crohn's disease; multiple sclerosis; infectious disease; cancer;	
KW		cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;	
KW		immune disorder; haematopoietic disorder; dyslipidaemia;	
KW		metabolic syndrome X; gene; ss.	
OS		Homo sapiens.	
XX			
PN		WO2003023001-A2.	
XX			
PD		20-MAR-2003.	
XX			
PF		09-SEP-2002; 2002WO-US028538.	

XX	07-SEP-2001;	2001US-0318120P.
PR	07-SEP-2001;	2001US-0318184P.
PR	10-SEP-2001;	2001US-0318430P.
PR	17-SEP-2001;	2001US-0322636P.
PR	17-SEP-2001;	2001US-0322781P.
PR	17-SEP-2001;	2001US-0322816P.
PR	17-SEP-2001;	2001US-0322817P.
PR	19-SEP-2001;	2001US-0323519P.
PR	20-SEP-2001;	2001US-0323631P.
PR	20-SEP-2001;	2001US-0323636P.
PR	25-SEP-2001;	2001US-0324969P.
PR	25-SEP-2001;	2001US-0325091P.
PR	26-SEP-2001;	2001US-0324990P.
PR	14-DEC-2001;	2001US-0341144P.
PR	26-FEB-2002;	2002US-0359599P.
PR	05-MAR-2002;	2002US-0361663P.
PR	03-MAY-2002;	2002US-0377908P.
PR	17-MAY-2002;	2002US-0381483P.
PR	29-MAY-2002;	2002US-0383863P.
PR	02-JUL-2002;	2002US-0393332P.
PR	17-JUL-2002;	2002US-0396412P.
PR	13-AUG-2002;	2002US-0403517P.
PR	06-SEP-2002;	2002US-00236417.

(CURA-) CURAGEN CORP.

PI Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangoli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khrantsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigar M, Patcucajan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twemlow N, Vernet CAM, Voss EZ;
PI Zernusen BD, Zhong M;

WPI; 2003-313241/30.
P-PSDB; ABR54263.

Novel human proteins and nucleic acid encoding the proteins, useful for diagnosis, treatment and prevention of disorders involving the human protein or nucleic acid e.g. cardiac and neurological disorders.

Claim 20; Page 243; 460pp; English.

The present invention describes isolated human NOVX proteins, where X is 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in ABR4167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant, hypotensive, dermatological, anorectic, immunosuppressive, cytostatic, antidiabetic, antiinfectility, haemostatic, antiinflammatory, anti-HIV, antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic, antiparkinsonian and antilipaeamic activities, and can be used in gene therapy. NOVX proteins are useful for treating or preventing a pathology associated with a NOVX protein in humans and for treating a syndrome associated with the human disease. NOVX nucleic acids, proteins and antibodies can be used in the treatment and diagnosis of cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, valvular disease, tuberosus sclerosis, scleroderma, obesity, transplantation, congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis, infectious disease, anorexia, cancer-associated cachexia, cancer, Alzheimer's disease, Parkinson's disease, immune disorders, haematopoietic disorders, dyslipidaemias, and metabolic syndrome X. ACC62346 to ACC62465 represent PCR primers and probes for human NOVX sequences, which are used in examples from the present invention.

CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Alignment Scores: 1.17e-249 Length: 1650
Pred. No.: 2480.00 Matches: 478
Score: 91.92% Conservative: 0
Percent Similarity: 91.92% Mismatches: 0
Best Local Similarity: 91.92% Indels: 42
Query Match: 98.77% Gaps: 1
DB: 7

US-10-622-516-2 (1-478) x ACC62332 (1-1650)

QY	1	MecProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal	20
DB	22	ATGCCCTGGATCACTTCTTGAATGCAGAGCTTGGCCAAAGATGTTGGGATTTGT	81
QY	21	AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp	40
DB	82	GCCCTTGAGATCTATTTCTCTTCAATATGTTGATCAAGCAGAGTTGAAAAATATGAT	141
QY	41	GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
DB	142	GGTGTAGATGCTGGGAAGTATACCATTTGGCTTGGCCAGGCCAAGATGGCTTGCACA	201
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn	80
DB	202	GATAGAGAAGATATTAACTCTCTTTCATGACTGTGGTTCAAGATCTTATGGAGAGAAAT	261
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
DB	262	AACCTTTCCTATGATTGCAITGGGCGGCTGGAAGTTGGAACAGACAAATCATCGACAAA	321
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
DB	322	TCAAGTCTGTGAAGACTAATTGTATGACAGCTGTTGAAGAGTCTGGGAATACAGATATA	381
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
DB	382	GAAAGAAATCGACACAACTAATGCATGCTATGGAGCAGACAGCTGTGTTCAATGCTGTT	441
QY	141	AsnTrpIleGluSerSerSerTrpAsp	149
DB	442	AACTGATTGAGTCCAGCTCTTGGAATGACGGTATGCCCTGTAGTTCAGAGAGATATT	501
QY	149		149
DB	502	GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGACAGTACTTGCCTA	561
QY	150		158
DB	562	ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGACACATATGCAACAT	621
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu	178
DB	622	GCCTATGATTTTACAAAGCTGATATGCTATCTGAATATCTTAATAGTAGATGGAAACTC	681
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
DB	682	TCCATACAGTGTACTACCTCAGTGCATTAGACCGCTGCTATTCTGTACTGCAAAAAAGATC	741
QY	199	HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
DB	742	CATGCCCAAGTGCAGAAAGAGGAATATTAAGATTTTACCTTGAATGATTTGGCTTC	801
QY	219	MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu	238
DB	802	ATGATCTTTCACCTCAACCATATTCCTAACTGTTCAAGAAATCTTAAGCTCGGATGTGCTG	861
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla	258
DB	862	AATGACTTCCTTAATGACCAAGAAATAGATATAAATAGTATCTAATAGTGCGCTGGAAGCC	921
QY	259	PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet	278
DB	922	TTTGGGATGTTAAATTAGAAAGACACTTATGATTAAGATGTGGAAGAAAGGCATTATATG	981

QY	279	LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln	298
DB	982	AAGGCTAGCTCTGAACCTCTTCACTCAGAAACAAAGCATCTTTACTGTATCAAAATCAA	1041
QY	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
DB	1042	AATGGAATATGTACACATCTTCAGTATATGTTCCCTTGCAATCTGTTCTAGACAGTAC	1101
QY	319	SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
DB	1102	TCACTCAGCAATTAGCAGGGAAGAAATTGAGTGTTCCTTAATGCTTCTGTTGGCT	1161
QY	339	AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys	358
DB	1162	GCCACTCTGTACTCTCTTAAAGTCACACAAAGATGCTACACCGGGTCTGCTTGTATAA	1221
QY	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
DB	1222	ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTGTGGCACA	1281
QY	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle	398
DB	1282	GATGTCTTCGCTGAAACATGAGCTCAGAGAGACACCATCATTTGGTCAACTATATT	1341
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu	418
DB	1342	CCCCAGGTTCAATGATTCACCTTTGAAGAACGTGTACTTAAAGGTGATGAA	1401
QY	419	LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly	438
DB	1402	AAGCAGAGAAGAACTTACGCTCGGCGTCCCACTCCAATGATGACACTTTGGATGAAGGA	1461
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal	458
DB	1462	GTAGACTTGTGCAATCAAAACATAGCACTGAGCATATTTCCAAGCCCTGCCAAGAAAGTA	1521
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
DB	1522	CCAAGACTCCCTGCACAGCAGCAGAACCTGAAGCAGCTGTCAATTAGTATGGGGAACAT	1581
RESULT 14			
ID	ADE38394	standard; DNA; 1650 BP.	
XX	AC	ADE38394;	
XX	AC		
DT	29-JAN-2004	(first entry)	
XX	DE	Human protein 9389 gene sequence.	
XX	KW	tumorigenic disorder; angiogenic disorder; aberrant gene expression;	
KW	KW	aberrant protein activity; cytostatic; antithyroid; antidiabetic;	
KW	KW	ophthalmological; cancer; breast cancer; colon cancer; lung cancer;	
KW	KW	prostatic cancer; Grave's disease; diabetic retinopathy; gene; ds;	
OS	XX		
OS	XX	Homo sapiens.	
XX	XX		
FH	Key	Location/Qualifiers	
FT	CDS	22..1584	
FT		/*tag= a	
FT		/product= "Human protein 9389"	
PN	WO2003065006-A2.		
XX	XX		
PD	07-AUG-2003.		
XX	XX		
PF	30-JAN-2003; 2003WO-US002588.		
XX	XX		
PR	31-JAN-2002; 2002US-0353600P.		
PR	15-MAR-2002; 2002US-0364517P.		
PR	09-APR-2002; 2002US-0371075P.		

PR 10-APR-2002; 2002US-0371507P.
PR 16-APR-2002; 2002US-0372984P.
PR 19-APR-2002; 2002US-0374194P.
PR 24-MAY-2002; 2002US-0382995P.
PR 31-MAY-2002; 2002US-0385023P.
PR 14-JUN-2002; 2002US-0388853P.
PR 17-JUN-2002; 2002US-0389395P.
PR 25-JUN-2002; 2002US-0391324P.
PR 15-JUL-2002; 2002US-0395944P.
PR 22-JUL-2002; 2002US-0397726P.
PR 13-AUG-2002; 2002US-0403046P.
PR 22-AUG-2002; 2002US-0405155P.
PR 27-AUG-2002; 2002US-0406361P.
PR 25-OCT-2002; 2002US-0421195P.
PR 12-NOV-2002; 2002US-0425456P.
PR 19-NOV-2002; 2002US-0427626P.
PR 10-DEC-2002; 2002US-0432122P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
XX Hunter JJ, Macbeth KJ, Tsai F, Lesoon A, Lightcap ES;
PI Williamson MW, Rudolph-Owen LA;
XX
XX WPI; 2003-646176/61.
DR P-PSDB; ADE38395.
XX
PT Treating subject having tumorigenic disorder or angiogenic disorder
PT caused by aberrant polypeptide e.g., N-formylpeptide receptor or nucleic
PT acid, by administering a modulator.
XX
XX
PS Disclosure; SEQ ID NO 55; 454bp; English.
XX
XX This invention relates to a novel method of treating a human subject
CC having a tumorigenic disorder or angiogenic disorder, caused by aberrant
CC gene expression or activity of an isolated protein, by administering a
CC modulator. The modulator may have cytostatic, antithyroid, antidiabetic
CC or ophthalmological activity. The method is useful for treating a subject
CC having a tumorigenic or angiogenic disorder, in particular for treating
CC cancer (for example breast cancer, colon cancer, lung cancer or prostatic
CC cancer) and, for example, Grave's disease and diabetic retinopathy. The
CC present sequence is a DNA sequence which encodes the novel isolated human
CC protein 9389 of the invention.
XX
XX Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 1.17e-249 Length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 9 Gaps: 1
US-10-622-516-2 (1-478) x ADE38394 (1-1650)
QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 22 ATGCTGATCACTTCTTGAATGCAAGAGCTTGCGCCAAAGATGTGGATTGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 82 GCCCTGAGATCTATTTCTTCAATATGTGTGATCAAGCAGAGTTGGAATAATATGAT 141
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 142 GGTGTAGATGCTGGAGATATACCATTTGGCTTGCGCCAGGCCAAGATGGCTTCTGCACA 201
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATTAACTCTTTGCATGACTGTGTTCCAGAACTTATGGAGAGAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100

Db 262 AACCTTTCCTATGATTGCATTGGCGGCTGGAAGTTGGAACAGACAATCATCGACAAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAAGCTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 382 GAAGGAATCGACACACTAATGCATGCTATGAGGCACAGCTGCTCTCAATGCTGTT 441
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 442 AACTGGAATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGCTAGTTGCAGAGATATT 501
QY 149 ----- 149
Db 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGGAGTTGGACAGTAGCTTGCTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATTGGGCCAAATGCTCCTTAAATTTTGAACGAGGGCTTGTGGACACATATGCAACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 622 GCCTATGATTTTACAGACCTGATATGCTATCTGAATATCTATAGTAGTGGAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 682 TCCATACAGTGTCACTACCTCAGTGCATTAGACCCGCTGCTATTCTACTGCAAAAAGATC 741
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCCAGTGGCAGAAAGGAGAAATGATAAAGATTTAACCTTGAATGATTTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 802 ATGATCTTTCACTCACCAATTTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCCTAATGACCGAATAGAGATAAAATAGTATCTATAGTGCGCTGGAAGCC 921
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 922 TTTGGGATGTTAAATTAGAACACACCTACTTTGATAGAGATGTGGAAGGCATTTTATG 981
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 982 AAGGCTAGCTCTGAACCTTTCACTCAGAAACAAAGGCACTTTACTGTATCAAAATCAA 1041
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1042 AATGAAATATGTACACACATCTTCAGTATATGTTCCCTGCATCTGTTCTAGCACAGTAC 1101
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1102 TCACCTCAGCAATTAGCAGGGAAGAATTTGAGTGTCTTATAGTCTGCTTGGCT 1161
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1162 GCCACTCTGTACTCTCTTAAAGTCAACAAGATGCTACACCGGGGTCTGCTTGATATAA 1221
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1222 ATTAACAGCAAGTTATGTATCTTAAATCAAGGCTTGATTCAAGAACTGTGTGGACCA 1281
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db 1282 GATGTCTTGTGCTGAAAAATGAAGCTCAGAGAGACACCATGATTTGGTCAACTATATT 1341
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAspGlu 418
Db 1342 CCCAGGGTTCATATAGATTCACTCTTTGAAGGAACGTGTACTTAGTTAGGTGATGAA 1401

Qy	419	LYSHISARGARGTHRITYRALAARGARGPROTHIPROASNAASPSPTHRLEUASPGUGLY	438
Db	1402	AAGCACAGAAAGAACTTACGCTCGCGCTCCCACTCCAATGATGACACTTTGGATGAAGGA	1461
Qy	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysIleVal	458
Db	1462	GTAGGACTTGTGCATTCAACACATAGCACTGAGCATATTCCAAAGCCCTGCCAAGAAAGTA	1521
Qy	459	ProArgLeuProAlaThrAlaIleAlaGluProGluIleAlaValIleSerAsnGlyGluHis	478
Db	1522	CCAAGACTCCCTGCCACAGCAGCAAGACCTGAAGCAGCTGTCTATTAGTAATGGGAACAT	1581
RESULT 15			
ADE76933			
ID	ADE76933	standard; cDNA; 3722 BP.	
XX	AC	ADE76933;	
XX	DT	29-JAN-2004 (first entry)	
XX	DE	Human cDNA differentially expressed in a liver disorder #70.	
XX	KM	human; ss; gene; liver disorder; hyperlipidaemia; hypertension;	
KM	KW	type II diabetes; tumour; liver; inflammatory disorder;	
KW	KM	immune response disorder; high-throughput screening;	
XX	OS	differential gene expression; gene therapy.	
XX	OS	Homo sapiens.	
XX	PN	US2003108871-A1.	
XX	PD	12-JUN-2003.	
XX	PF	30-JUL-2001; 2001US-00919039.	
XX	PR	28-JUL-2000; 2000US-0222113P.	
PA		(KASE/) KASER M R.	
PI		Kaser MR;	
XX	DR	WPI; 2004-031227/03.	
XX	PT	Composition comprising several cDNAs that are differentially expressed in	
PT	PT	treated human C3A liver cell cultures, useful for treating liver	
PT		disorders.	
XX	PS	Claim 1; SEQ ID NO 98; 41pp; English.	
XX		The invention relates to a composition comprising several cDNAs that are	
CC		differentially expressed in a liver disorder. The composition is useful	
CC		for treating liver disorder such as hyperlipidaemia, hypertension, type	
CC		II diabetes, tumours of the liver and disorders of the inflammatory and	
CC		immune response. The composition is useful for a high-throughput method	
CC		of screening several molecules or compounds to identify a ligand which	
CC		specifically binds a cDNA. A protein encoded by the cDNA is useful for a	
CC		high-throughput method for using a protein to screen several molecules or	
CC		compounds to identify at least one ligand which specifically binds the	
CC		protein which involves combining the protein encoded by the cDNA with	
CC		several of molecules or compounds under conditions to allow specific	
CC		binding, and detecting specific binding between the protein and a	
CC		molecule or compound, therefore identifying a ligand which specifically	
CC		binds the protein. The composition is useful for detecting and	
CC		quantifying differential gene expression, can be used in gene therapy, to	
CC		formulate prognosis and to design a treatment regimen and to monitor the	
CC		efficacy of treatment. The present sequence represents a cDNA	
CC		differentially expressed in a liver disorder.	
XX	SEQ	Sequence 3722 BP; 1121 A; 645 C; 860 G; 1094 T; 0 U; 2 Other;	

Alignment Scores:

3.82e-249

Length:

3722

Score:	2480.00	Matches:	478
Percent Similarity:	91.92%	Conservative:	0
Best Local Similarity:	91.92%	Mismatches:	0
Query Match:	98.77%	Indels:	42
DB:	10	Gaps:	1

US-10-622-516-2 (1-478) x ADE76933 (1-3722)

QY	1	MePProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal	20
Db	249	ATGCTGGATCACCCTCTTGAATGCAGAACTTGCTGGCCAAAAGATGTGGAAATTGTT	308
QY	21	AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp	40
Db	309	GCCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAATAATATGAT	368
QY	41	GlyValaAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
Db	369	GGTGTAGATGCTGGAAGTATACCATTTGGCTTGGCCAGGCCAAGATGGGCTTCTGCACA	428
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn	80
Db	429	GATAGAGAAGATATTAACTCTCTTTGCATGACGTGTGGTTCGAAATCTTATGAGAGAAAT	488
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
Db	489	AACCTTCTATGATTTGATTTGGCGCGCTGGAAGTTGGAACAGACAATCATCGACAAA	548
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
Db	549	TCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA	608
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
Db	609	GAAAGAAATGCACACAACTAATGCATGCTATGAGGACACAGCTGCTGCTTCAATGCTGTT	668
QY	141	AsnTrpIleGluSerSerSerTrpAsp	149
Db	669	AACTGATTCAGTCCAGCTCTTGGAGTGCACGCTATGCCCTGTAGTTGCAGAGATATT	728
QY	149		149
Db	729	GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGACGAGTAGCTCTGCTA	788
QY	150		158
Db	789	ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT	848
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysIleu	178
Db	849	GCCTATGATTTTCAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAAAACTC	908
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
Db	909	TCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTAATGCAAAAAAGATC	968
QY	199	HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
Db	969	CATGCCCAAGTGGCAGAAAGAGGGAATGATAAAGAATTTTACCTTGAATGATTTTGGCTTC	1028
QY	219	MetIlePheHisSerProTyrCysLysIleuValGlnLysSerLeuAlaArgMetLeuLeu	238
Db	1029	ATGATCTTTCACTCAACCATATTGTAACTGGTTCAAGAAATCTTACCTCGGATGTTGCTG	1088
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla	258
Db	1089	AATGACTTCTTAAATGACAGAAATAGAGATAAAATAATGATATCTATAGTGCGCTGGAAGCC	1148
QY	259	PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet	278
Db	1149	TTTGGGAGATGTTAAATTAGAAAGACACCTACTTTGATAGAGATGTGGAAGCAATTTATG	1208
QY	279	LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln	298

Db	1209		1268
		AAGCTAGCTCTGAACCTCTTCAGTCAGAAAACAAAGGCATCTTACTTGATCAAAATCAA	
QY	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
Db	1269		
		AATGGAATATGTACACATCTTCAGTATATGGTCCCTGCATCTGTTCTAGCACAGTAC	1328
QY	319	SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
Db	1329		
		TCACTCAGCAATTAGCAGGGAGAAGAAATTGAGTGTCTTATGGTCTGCTTGGCT	1388
QY	339	AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys	358
Db	1389		
		GCCACTCTGACTCTCTTAAGTCACACAAAGATGCTACCGGGGCTGCTCTTGATPAA	1448
QY	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
Db	1449		
		ATAACAGCAAGTTATGTGATCTTAATCAAGCTTGATCAAGAACTGGTGTGGCAACA	1508
QY	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisLeuValAsnTyrIle	398
Db	1509		
		GATGCTCTCGCTGAACAACATGAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT	1568
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrPyrLeuValArgValAspGlu	418
Db	1569		
		CCCCAGGGTTCAATAGATTCACTTTTGAAGAACGTGTACTTAGTTAGGGTGGATGAA	1628
QY	419	LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly	438
Db	1629		
		AAGCACAGAAAGAACTTACGCTGCGCTCCCACTCCAAATGATGACACTTTGGATGAAGCA	1688
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal	458
Db	1689		
		GTAAGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA	1748
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGlnHis	478
Db	1749		
		CCAAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAGCTGTCAATTAATGGGGAACAT	1808

Search completed: June 24, 2004, 13:53:21
Job time : 541 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 24, 2004, 15:10:46 ; Search time 524 Seconds
(without alignments)
4178.933 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPEGLPLNNAEACWPKDVGIV.....PRLPATPAEPEAAVISNGEH 478

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3017426 segs, 2290544650 residues
Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10622516/runat_23062004_162636_775/app_query.fasta_1.647
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10622516_@CGN_1_1_723_@runat_23062004_162636_775
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2: *
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq: *
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq: *
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq: *
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	2511	100.0	2002	14	US-10-193-295-1	Sequence 1, Appli
2	2511	100.0	2002	16	US-10-622-516-1	Sequence 1, Appli
3	2480	98.8	1564	13	US-10-236-417-203	Sequence 203, App
4	2480	98.8	1564	13	US-10-236-417-207	Sequence 207, App
5	2480	98.8	1564	13	US-10-307-817-369	Sequence 369, App
6	2480	98.8	1564	13	US-10-307-817-373	Sequence 373, App
7	2480	98.8	1601	13	US-10-236-417-191	Sequence 191, App
8	2480	98.8	1601	13	US-10-307-817-363	Sequence 363, App
9	2480	98.8	1650	13	US-10-236-417-181	Sequence 181, App
10	2480	98.8	1650	13	US-10-236-417-183	Sequence 183, App
11	2480	98.8	1650	13	US-10-236-417-185	Sequence 185, App
12	2480	98.8	1650	13	US-10-236-417-189	Sequence 189, App
13	2480	98.8	1650	13	US-10-236-417-193	Sequence 193, App
14	2480	98.8	1650	13	US-10-236-417-197	Sequence 197, App
15	2480	98.8	1650	13	US-10-236-417-201	Sequence 201, App
16	2480	98.8	1650	13	US-10-236-417-205	Sequence 205, App
17	2480	98.8	1650	13	US-10-307-817-359	Sequence 359, App
18	2480	98.8	1650	13	US-10-307-817-371	Sequence 371, App
19	2480	98.8	1650	13	US-10-307-817-445	Sequence 445, App
20	2480	98.8	1650	15	US-10-354-358-55	Sequence 55, Appl
21	2480	98.8	3722	10	US-09-919-039-98	Sequence 98, Appl
22	2475	98.6	1593	13	US-10-236-417-187	Sequence 187, App
23	2475	98.6	1593	13	US-10-307-817-361	Sequence 361, App
24	2475	98.6	1608	13	US-10-236-417-195	Sequence 195, App
25	2475	98.6	1608	13	US-10-236-417-211	Sequence 211, App
26	2475	98.6	1608	13	US-10-307-817-365	Sequence 365, App
27	2475	98.6	1608	13	US-10-307-817-375	Sequence 375, App
28	2475	98.6	1612	13	US-10-236-417-199	Sequence 199, App
29	2475	98.6	1612	13	US-10-236-417-209	Sequence 209, App
30	2475	98.6	1612	13	US-10-307-817-357	Sequence 357, App
31	2475	98.6	1612	13	US-10-307-817-367	Sequence 367, App
32	2362	94.1	3275	13	US-10-205-331-15	Sequence 15, Appl
33	2362	94.1	3275	16	US-10-388-934-37	Sequence 37, Appl
34	1691	67.3	170834	9	US-09-835-232-7	Sequence 7, Appli
35	1691	67.3	170834	15	US-10-308-485-7	Sequence 7, Appli
36	1631	65.0	3008	15	US-10-106-698-1926	Sequence 1926, Ap
37	1513	60.3	2058	9	US-09-969-708-417	Sequence 417, App
38	1513	60.3	2058	9	US-09-880-107-3810	Sequence 3810, Ap
39	1513	60.3	2058	13	US-10-240-425-1557	Sequence 1557, Ap
40	1513	60.3	2058	13	US-10-342-887-1161	Sequence 1161, Ap
41	1513	60.3	2058	13	US-10-172-118-1161	Sequence 1161, Ap
42	1513	60.3	2058	15	US-10-205-823-173	Sequence 173, App
43	1513	60.3	2058	15	US-10-177-293-207	Sequence 207, App
44	1513	60.3	2261	15	US-10-198-846-10322	Sequence 10322, A
45	1513	60.3	2377	15	US-10-252-157-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-10-193-295-1
; Sequence 1, Application US/10193295
; Publication No. US20020173018A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: C1001195DIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Human
US-10-193-295-1
Alignment Scores:

Pred. No.: 2.39e-293 Length: 2002
Score: 2511.00 Matches: 478
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0
US-10-622-516-2 (1-478) x US-10-193-295-1 (1-2002)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
DB 145 ATGCCTGATCACTTCCTTGAATGCAGAAAGCTTGCGCCAAAAGATGTGGAAATTGTT 204
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
DB 205 GCCCTTGAGATCTATTTCTCTCAATATGTGATCAAGCAGAGTTGGAAAATATGAT 264
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
DB 265 GGTGTAGATGCTGGAAGATATACCATTTGGCTGGGCCAGGCCAAGATGGGCTTGCCACA 324
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValAlaGlnAsnLeuMetGluArgAsn 80
DB 325 GATAGAGAAGATTAATCACTCTCTTTGCATGACTGTGTTCAGAATCTTATGAGAGAAAT 384
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
DB 385 AACCTTCCATGATTCATTTGGGGCGGCTGGAGTTGAACAGACAAATCATGCACAAA 444
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
DB 445 TCAAAAGTCTGTGAAGACTTAATTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA 504
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
DB 505 GAAGGAATCGACACAACATAATGCATGCTATGGAGCACAAGCTGCTCTTCAATGCTGTT 564
QY 141 AsnTrpIleGluSerSerSerTrpAspGlyLeuArgGlyThrHisMetGlnHisAlaTyr 160
DB 565 AACTGGATGAGTCCAGCTCTGGGATGGGCTCGTGGGACACATATGCAACATGCCCTAT 624
QY 161 AspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysIleSerIle 180
DB 625 GATTTTACAGACCTGATATGCTATCTGAATATCCTATAGTAGATGGAATAACTCTCCATA 684
QY 181 GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAla 200
DB 685 CAGTGTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACTGCAAAAAGATCCATGCC 744
QY 201 GlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIle 220
DB 745 CAGTGGCAGAAAAGAGGAAATGATTAAGATTTTACCTGAATGATTTTGGCTTCATGATC 804
QY 221 PheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp 240
DB 805 TTTCACCTACCACTAATTGTAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTGAATGAC 864
QY 241 PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGly 260
DB 865 TTCCTTAATGACAGATAGATAAAAATAGTATCTATAGTGCCCTGGAAGCCTTTGGG 924
QY 261 AspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAla 280
DB 925 GATGTTAAATAGAAAGACACCTAATTGTATAGATGTGAGAAAGCATTTATGAAGGCT 984
QY 281 SerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGly 300
DB 985 AGCTCTGAACCTCTTCAGTCAGAAAACAAAGGCATCTTACTGTATCAAAATCAAAATGA 1044
QY 301 AsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerPro 320
DB 1045 AATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTGCTAGCACAGTACTCACCT 1104

QY 321 GlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaIleThr 340
DB 1105 CAGCAATTAGCAGGGAAGAAATTGAGTGTCTTATGTTGTTGCTTGCCCACT 1164
QY 341 LeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThr 360
DB 1165 CTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTCTGCTCTTGATAAATAACA 1224
QY 361 AlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAspVal 380
DB 1225 GCAAGTTTATGTGATCTTAATCAAGCTTGATTCAAGAACTGGTGTGCACAGATGTC 1284
QY 381 PheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIleProGln 400
DB 1285 TTGCTGAAAACATGAAGCTCAGAGAGACACCCATCATTTGTCAACTATATCCCCAG 1344
QY 401 GlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGluLysHis 420
DB 1345 GGTTCATAGATTCACTCTTTGAAGGAACGTGTACTTAGTTAGGGTGATGAAGACAC 1404
QY 421 ArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGlyValGly 440
DB 1405 AGAAGACTTACGCTCGCGCTCCCACTCCAATGATGACACTTTGGATGAAGAGTAGGA 1464
QY 441 LeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysValProArg 460
DB 1465 CTGTGCATTCAAACATAGCAACTGAGCATATTTCCAAGCCCTGCCAAGAAAGTACCAAG 1524
QY 461 LeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
DB 1525 CTCCTGCCACAGCAGACAGAACTGAAGCAGCTGTCTATTAGTATGGGAACAT 1578

RESULT 2
US-10-622-516-1
; Sequence 1, Application US/10622516
; Publication No. US20040018545A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: THEREOF
; FILE REFERENCE: CL001195DIV2
; CURRENT APPLICATION NUMBER: US/10/622, 516
; PRIOR FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 10/193, 295
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/819, 993
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Human
US-10-622-516-1

Alignment Scores:
Pred. No.: 2.39e-293 Length: 2002
Score: 2511.00 Matches: 478
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0

US-10-622-516-2 (1-478) x US-10-622-516-1 (1-2002)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
DB 145 ATGCCTGATCACTTCCTTGAATGCAGAAAGCTTGCGCCAAAAGATGTGGAAATTGTT 204
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
DB 205 GCCCTTGAGATCTAATTTCTCTCAATATGTGATCAAGCAGAGTTGGAATAATATGAT 264

QY	41	GLYValAspAlaGlyLysThrTrpIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
Db	265	GGTGTAGATGCTGGAAAGTATACCATTTGGCTTGGGCCAGCCAAAGATGGGCTTCTGCACA	324
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn	80
Db	325	GATAGAGAATATTAACCTCTCTTGGCATGACTGTGGTTCAGAACTTATGAGAGAAAT	384
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
Db	385	AACCTTTCCTATGATTCATTTGGCGGGCTGGAGTTGGACACAGACAAATCATTCGACAAA	444
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
Db	445	TCAAAGCTGTGAAGACTTAATTGATGCAGCTGTTTGAGAGAGTCTGGGAATACAGATATA	504
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
Db	505	GAAAGAATCGACACAACTAATGCATGCTATGAGGACACAGCTGCTTCAATGCTGTT	564
QY	141	AsnTrpIleGluSerSerSerTrpAspGlyLeuArgGlyThrHisMetGlnHisAlaTyr	160
Db	565	AACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGTGGGACATATGCAACATGCTTAT	624
QY	161	AspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIle	180
Db	625	GATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGAAAACTCTCCATA	684
QY	181	GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIleHisAla	200
Db	685	CAGTGTACCTCAGTGCATTAGACCGCTGCTACTGTGTACTGCACAAAAGATCCATGCC	744
QY	201	GlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIle	220
Db	745	CAGTGCAGAAAGAGGAAATGATAAAGATTTTACCCTGAATGATTTGGCTTCATGATC	804
QY	221	PheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp	240
Db	805	TTTCACTCACCATATTTGTAACTGGTTCAGAAATCTCTAGCTCCGATGTTGCTGAATGAC	864
QY	241	PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGly	260
Db	865	TTCTTAATGACCAGATAGAGATAAAATAGTATCTAATGTGCCCTGGAAAGCTTTGGG	924
QY	261	AspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAla	280
Db	925	GATGTTAAATTAGAGACACCTACTTGTATAGAGATGTGGAGAGCATTTATGAAGGCT	984
QY	281	SerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGly	300
Db	985	AGCTCTGAACCTTTCAGTACAGAAACAAAGGCATCTTTACTTGTATCAAAATCAAAATGGA	1044
QY	301	AsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerPro	320
Db	1045	AATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTGTTCTAGCACAGTACTACCT	1104
QY	321	GlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThr	340
Db	1105	CAGCAATTAGCAGGGAAGAGAAATTGAGAGTGTCTTCTATGCTTCTGCTGGCTGCACCT	1164
QY	341	LeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThr	360
Db	1165	CTGTACTCTCTTAAAGTCACACAGATGCTACACCGGGTCTGCTTGATAAAATAACA	1224
QY	361	AlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPheAspVal	380
Db	1225	GCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGTGTGGCACCAAGATGTC	1284
QY	381	PheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIleProGln	400
Db	1285	TTGCTGAAAACATGAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATTTCCCCAG	1344

QY	401 GlySerIleAspSerLeuPheGluGlyThrTyrLeuValArgValAspGluLysHis 	420
Dd	1345 GGTCAATAGATTCACTTTGAAGAACGTGTACTTAGGGTGATGAAAGCAC 	1404
QY	421 ArgArgThrTyraAlaargArgProthrProAsnAspSPTHrLeuaspgluglYvalgily 	440
Dd	1405 AGAAGAACTTACGCTCGGCCTCCCACTCCAATGATGACACTTGATGAAGGAGTAAGA 	1464
QY	441 LeuValHissEraSnileAlathrgluHisileProSerProAlalyslysValProArg 	460
Dd	1465 CTGTGCATTCAAACATAAGCAACTGAGCATATTCCAAGCCTGCCAAGAAAGTACCAGA 	1524
QY	461 leuproAlathralaalagluprogluAlalaValilesEraSnglygluHis 	478
Dd	1525 CTCCTGCACAGCAGCAAGAACCTGAAGCAGCTGTTCATTAGTAATGGGAACAT 	1578

```

RESULT 3
US-10-236-417-203
; Sequence 203, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 203
; LENGTH: 1564
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(1562)
US-10-236-417-203

Alignment Scores:
Pred. No.:          9,17e-290          Length:          1564
Score:              2480.00             Matches:          478
Percent Similarity: 91.92%              Conservative:     0
Best Local Similarity: 91.92%            Mismatches:      0
Query Match:        98.77%              Indels:          42
DB:                  13                  Gaps:            1

US-10-622-516-2 (1-478) x US-10-236-417-203 (1-1564)

QY      1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
        |||||||
Db      2 ATGCGCTGATCACTTCCTTGATGACAGAGCTTGCTGGCCAAAGATGTGGGAATTGTT 61

```

Db	62	GCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCAGAGAGTGGAAAAATATGAT	121
QY	41	GLYValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
Db	122	GGTGTAGATGCTGGAAGATATACCATGGCTTGGGCCAGGCCAAGATGGGCTTCTGCACA	181
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn	80
Db	182	GATGAGAAAGATATTAACTCTCTTGGCATGACTGTGGTTCAGAACTTATGGAGAGAAAT	241
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
Db	242	AACCTTTCCTATGATTGCATTGGCGGCTGGAAAGTTGGAAACAGACAAATCATCGACAAA	301
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
Db	302	TCAAAGTCTGTGAAGACTAATTGTATGCAGCTGTTGAAGAGTCTGGGAAATACAGATATA	361
QY	121	GLUGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
Db	362	GAAGGAATCGACACACTAATGCATGCTATGGAGGCACAGCTGCTGTTCAATGCTGTT	421
QY	141	AsnTrpIleGluSerSerSerTrpAsp-----	149
Db	422	AACTGATTGAGTCCAGCTCTTGAGATGACGGTATGCCCTGGTAGTTCAGAGAGATATT	481
QY	149	-----	149
Db	482	GCTGTATTATGCCACAGGAATGCTAGACCTACAGGTGAGATTGGAGCAGTACTGTGCTA	541
QY	150	-----GlyLeuArgGlyThrHisMetGlnHis	158
Db	542	ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT	601
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu	178
Db	602	GCCATATGATTTTACAAAGCTGATATGCTATCTGAATATCCTATAGTAGTGAAGAACTC	661
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
Db	662	TCCATACAGTGTACTACCTGCTCATTAAGCCGCTGCTACTCTGCTACTGCAAAAAAGATC	721
QY	199	HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
Db	722	CATGCCCACTGGCAGAAAGGGAATGATAAAGATTTTACCTTGAATGATTTGGCTTG	781
QY	219	MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeu	238
Db	782	ATGATCTTTCACTCACCATATTGTAAACGTGGTTCAGAAATCTCTAGCTCGCATGTTGCTG	841
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla	258
Db	842	AATGACTTCCTTAATGACCAGAAATAGAGATAAAATAGTATCTATAGTGGCCTGGAAGCC	901
QY	259	PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet	278
Db	902	TTTGGGATGTTAAATTAGAAAGACACCTACTTTGATAGAGATGTGAGAAAGCATTTATG	961
QY	279	LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln	298
Db	962	AAGGCTAGCTCTGAACCTTTCAGTACAGAAAACAAAGGCATCTTTACTTGATCAAAATCAA	1021
QY	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
Db	1022	AATGAAATATGTACACATCTTCAGTATATGGTTCCTTGCATCTGTTCTAGCACAGTAC	1081
QY	319	SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
Db	1082	TCACCTCAGCAATTAGCAGGGAAGAAATTGAGTGTTCCTTATGTTCTGCTTGGCT	1141
QY	339	AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys	358

Db	1142	GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTCTGCTTGTATAA	1201
QY	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
Db	1202	ATAACGCAAGTTTATGATCTTAAATCAAGGCTTGATTCAGAACTGTGTGCACCA	1261
QY	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle	398
Db	1262	GATGCTTCGCTGAAAACATGAAGCTCAGAGAGACACCACCATCTTGGTCAACTATATT	1321
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAspGlu	418
Db	1322	CCCCAGGGTTCATAGATTTCACCTCTTGAAGAACGTTGTACTTAGTTAGGCTGATGAA	1381
QY	419	LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGly	438
Db	1382	AAGCACAGAAAGACTTACGCTCGGCGTCCACTCCAAATGATGACACTTTGGATGAAGGA	1441
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal	458
Db	1442	GTAGACTTGTGCATTCAACATAGCACTGAGCATATTCGAAGCCCTGCCAAGAAAGTA	1501
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
Db	1502	CCAAGACTCCCTGCACAGCAGACAGAACCTGAAGCAGCTGTCTATTAGTAATGGGAACAT	1561
RESULT 4			
US-10-236-417-207			
; Sequence 207, Application US/10236417			
; Publication No. US20040048256A1			
; GENERAL INFORMATION:			
; APPLICANT: Agree et al.			
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME			
; FILE REFERENCE: 21402-442C			
; CURRENT APPLICATION NUMBER: US/10/236,417			
; PRIOR FILING DATE: 2003-01-06			
; PRIOR APPLICATION NUMBER: US60/318,120			
; PRIOR FILING DATE: 2001-09-01			
; PRIOR APPLICATION NUMBER: US60/318,430			
; PRIOR FILING DATE: 2001-09-10			
; PRIOR APPLICATION NUMBER: US60/322,781			
; PRIOR FILING DATE: 2001-09-17			
; PRIOR APPLICATION NUMBER: US60/318,184			
; PRIOR FILING DATE: 2001-09-07			
; PRIOR APPLICATION NUMBER: US60/361,663			
; PRIOR FILING DATE: 2002-03-05			
; PRIOR APPLICATION NUMBER: US60/396,412			
; PRIOR FILING DATE: 2002-07-17			
; PRIOR APPLICATION NUMBER: US60/322,636			
; PRIOR FILING DATE: 2001-09-17			
; PRIOR APPLICATION NUMBER: US60/322,817			
; PRIOR FILING DATE: 2001-09-17			
; PRIOR APPLICATION NUMBER: US60/322,816			
; PRIOR FILING DATE: 2001-09-17			
; PRIOR APPLICATION NUMBER: US60/323,519			
; PRIOR FILING DATE: 2001-09-19			
; Remaining Prior Application data removed - See File Wrapper or PALM.			
; NUMBER OF SEQ ID NOS: 341			
; SOFTWARE: Custom			
; SEQ ID NO 207			
; LENGTH: 1564			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (2)..(1562)			
US-10-236-417-207			
Alignment Scores:			
Pred. No.: 9.17e-290			
Score: 2480.00			
Percent Similarity: 91.92%			
Best local Similarity: 91.92%			
Length: 1564			
Matches: 478			
Conservative: 0			
Mismatches: 0			

Query Match: 98.77% Indels: 42
DB: 13 Gaps: 1
US-10-622-516-2 (1-478) x US-10-236-417-207 (1-1564)

QY	1	MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal	20
DB	2	ATGCTTGATCACTTCCTTTGAATGCAGAAAGCTTGCTGGCCAAAGATGGGAATTGTT	61
QY	21	AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp	40
DB	62	GCCCTTGAGATCTATTTCTCTCAATATGTGATCAAGCAGAGTTGGAATAATATGAT	121
QY	41	GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
DB	122	GGTGTAGATGCTGGAAGATATACCATGGCTTGCCAGGCCAAGATGGGCTTCTGCACA	181
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn	80
DB	182	GATAGAGAAGATATTAACCTCTTTGCATGACTGTGGTTCAGAACTTATGAGAGAAT	241
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
DB	242	AACCTTTCTATGATTGCAATTGGCGGCTGGAAGTTGGAACAGACAAATCATGCACAA	301
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
DB	302	TCAAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGAATACAGATATA	361
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
DB	362	GAAGGAATGCACAACTAATGCATGCTATGAGGACACAGCTGCTGTTCAATGCTGTT	421
QY	141	AsnTrpIleGluSerSerSerTrpAsp-----	149
DB	422	AACTGGATTGATCCAGCTCTTGGGATGACGCTATGCCCTGTAGTTGCAGAGATATT	481
QY	149	-----	149
DB	482	GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGACAGCAGTAGCTTGCTA	541
QY	150	-----GlyLeuArgGlyThrHisMetGlnHis	158
DB	542	ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGCTTCGTGGACACATATGCAACAT	601
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu	178
DB	602	GCCTATGATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGATGAGAAATCTC	661
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
DB	662	TCCATACAGTGTACTACCTCAGTGCATTAGACCGCTGCTACTCTGCTACTGCAGAAAAGATC	721
QY	199	HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
DB	722	CATGCCCACTGGCAGAAAGGGAATGATAAAGATTTTACCTTGAATGATTTGGGCTTC	781
QY	219	MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu	238
DB	782	ATGATCTTTCACTCACCATATTTGTAACCTGCTTCAGAAATCTCTAGCTCGGATGTGCTG	841
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla	258
DB	842	AATGACTTCCTTAATGACCAGAAATAGATAAATAATAGTATCTATAGTGGCCTGGAACC	901
QY	259	PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet	278
DB	902	TTTGGGGATGTTAAATTAGAGACACTACTTTGATAGAGATGCGAAGGCAATTTATG	961
QY	279	LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln	298
DB	962	AAGGCTAGCTCTGAACCTTTCAGTCAGAAAAACAAGGCATTTTACTTGTATCAAAATCAA	1021

QY	299	AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr	318
DB	1022	AATGGAATATGTACACACATCTTCAGTATATGTTCCCTTGCACTCTGTCTAGCACAGTAC	1081
QY	319	SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla	338
DB	1082	TCACCTCAGCAATTAGCAGGGAAGAATTTGAGTGTCTTCTTATGTTCTGTTGGCT	1141
QY	339	AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys	358
DB	1142	GCCACTGTACTCTCTTAAGTTCACACAGATGTACACCGGGCTGCTCTTGATATAA	1201
QY	359	IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro	378
DB	1202	ATAACAGCAAGTTATGTGATCTTAATCAAGCTTGATTCAAGAACTGTGTGGCACCA	1261
QY	379	AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisLysLeuValAsnTyrIle	398
DB	1262	GATGTCTTCGCTGAAAACATGAGCTCAGAGAGACACCATTCATTGTCCAACTATATT	1321
QY	399	ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu	418
DB	1322	CCCAGGGTTCATATAGATTCACTTTTGAAGGAACGTGTACTTAGTGGTGATGAA	1381
QY	419	LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly	438
DB	1382	AAGCACAGAAGACTTACGCTCGCGCTCCCACTCCAATGATGACACTTTGGATGAAGGA	1441
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal	458
DB	1442	GTAGGACTTGTGCATTCAACATTAGCAACTGAGCATATTCCAAAGCCCTGCCAAGAAAGTA	1501
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
DB	1502	CCAAGACTCCCTGCACAGCAGCAAGAACTGAAGCAGCTGTCAATTAGTATGGGAACAT	1561

RESULT 5

US-10-307-817-369
; Sequence 369, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307, 817
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 369
; LENGTH: 1564
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(1561)
US-10-307-817-369

Alignment Scores:
Pred. No.: 9.17e-290 Length: 1564
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 13 Gaps: 1

US-10-622-516-2 (1-478) x US-10-307-817-369 (1-1564)	
QY	1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db	2 ATGCTTGATCACTTCCTTTGAATGCAGAAAGCTTGCCAAAGATGTGGGAATTGTT 61
QY	21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40

Db 62 GCCCTTGAGATCTATTTTCTCTCTCAATATGTTGATCAAGCAGAGTTGCAAAAATATGAT 121
QY 41 GlyValAspAlaGlyLysThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 122 GGTGTAGATGCTGGAAAGTATACCATTTGGCTTGGCCAGGCCAAGATGGGCTTCTGCACA 181
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 182 GATAGAGAAGATATTAACTCTCTTTCATGACTGTGGTTCAGAATCTTATGAGAGAAAT 241
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 242 AACCTTTCCTATGATTGCATTTGGCGGCTGGAAAGTTGGAACAGAGCAATCATCGACAAA 301
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 302 TCAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAAGAGTCTGGGAATACAGATATA 361
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 362 GAAGGAATCGACACAACTAATGATGATGAGGACAGCTGTCTTCAATGCTGTT 421
QY 141 AsnTrpIleGluSerSerSerTyrAsp----- 149
Db 422 AACTGATGTAGTCCAGCTCTTGAGATGACGGTATGCCCTGTAGTTCAAGAGATATT 481
QY 149 ----- 149
Db 482 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGGAGTTGGAGCAGTAGCTCTGCTA 541
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 542 ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGACACACATATGCAACAT 601
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 602 GCCTATGATTTTACAGCCTGATATGCTAATCTGAATATCCTATAGTAGATGAAAAAATC 661
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 662 TCCATACAGTGTCTACCTCAGTCATGACATGACCGCTGCTACTCTGTACTGCAAAAAAGATC 721
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 722 CATGCCCACTGGCAGAAAGGGAAATGATTAAGATTTAACCTTGAATGATTTTGGCTTC 781
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 782 ATGATCTTTCACCTCACCATTAATTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTG 841
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 842 AATGACTTCTTAATGACCAAGATAGAGATAAATACTATCTATAGTGGCCTGGAAGCC 901
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 902 TTTGGGATGTTAAATTAGAGACACCTACTTGATAGAGATGTGAGAAAGCATTTATG 961
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 962 AAGGCTAGCTGAACTCTTCAGTCAGAAAACAAGCATCTTACTTGTATCAAAATCAA 1021
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1022 AATGAAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTCTAGCACAGTAC 1081
QY 319 SerProGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1082 TCACCTCAGCAATTAGCAGGGAAGAAATGGAGTGTTCCTATATGTTCTGTTGGCT 1141
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1142 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTGCTTTGATAAA 1201

QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1202 ATAACAGCAAGTATTATGTATCTTAATCAAGGCTTGATTCAGAACTGCTGCGACCA 1261
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisIleValAsnTyrIle 398
Db 1262 GATGCTTCGCTGAAAACATGAAAGCTCAGAGAGACACCACATCATTTGGTCAACTATATT 1321
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
Db 1322 CCCAGGGTTCAATGATTCATCTCTTGAAGAAACGTGTACTTAGTTAGGGTGATGAA 1381
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGlu 438
Db 1382 AAGCAGAGAAGAACTTACGCTCGCGCTCCCACTCCAATGATGACACTTTGGATGAAGGA 1441
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db 1442 GTAGACTTGTGCATTCAAACATAGCAACTGAGCATATTTCCAAGCCCTGCCAAGAAAGTA 1501
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db 1502 CCAAGACTCCCTGCCACAGACAGAACTTGAAGCAGCTGTCTATTAATGGGGAACAT 1561

RESULT 6
US-10-307-817-373
; Sequence 373, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Agree et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307, 817
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 373
; LENGTH: 1564
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (2)..(1561)
US-10-307-817-373

Alignment Scores:
Pred. No.: 9.17e-290 Length: 1564
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 13 Gaps: 1

US-10-622-516-2 (1-478) x US-10-307-817-373 (1-1564)
QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 2 ATGCCTGATCACTTCTTGAATGCAAGAGCTTGCTGGCCAAAAGATGGGAATTTGTT 61
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 62 GCCCTTGAGATCTAATTTCTCTTCTCAATATGTGATCAAGCAGAGTTGGAATAATATGAT 121
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 122 GGTGTAGATGCTGGAAAGTATACCATTTGGCTTGGCCAGGCCAAGATGGGCTTCTGCACA 181
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 182 GATAGAGAAGATATTAACTCTCTTTCATGACTGTGGTTCAGAATCTTATGAGAGAAAT 241
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100

Db 242 AACCTTTCCTATGATGTGCAATGGGCGGCTGGAAGTTGGAACAGAGACATCATCGACAAA 301
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 302 TCAAGTCTGTGAGACTAATTGTATGCAGCTGTTGGAAGAGCTCGGAATACAGATATA 361
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 362 GAAAGAAATGCACACAACATAATGCATGCTATGGAGGACACAGCTGCTCTTCAATGCTGTT 421
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 422 AACTGGATTGAGTCCAGCTCTGGGATGGACGGTATGCCCTGGTAGTTGCAGAGATATT 481
QY 149 ----- 149
Db 482 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGAGCAGTAGCTTGCTA 541
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 542 ATTGGGCCAATGCTCCTTAATTTTTGAACGAGGGGCTGCTGGGACACATATGCACAT 601
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGlyTyrProIleValAspGlyLysLeu 178
Db 602 GCCTATGATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAATACTC 661
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 662 TTCATACAGTGTACCTCAGTGCATTAGACCGCTGCTACTGCTCTACTGCCAAAAGATC 721
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 722 CATGCCAGTGGCAGAAAGGGAATGATAAAGATTTTACCTTGAATGATTTGGCTTC 781
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeu 238
Db 782 ATGATCTTTCACTCACCATATTGTAACCTGTTCAGAAATCTTAGCTCGGATGTTGCTG 841
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 842 AATGACTTCCTTAATGACCAAGATAAGATAAAAATAGTATCTATAGTGGCCCTGGAAGCC 901
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 902 TTGGGGATGTTAAATTAGAGACACCTACTTGTATAGATGTGAGAAAGGCATTATTATG 961
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 962 AAGGTAAGCTTGAACTCTTCAGTCAGAAAACAAAGGCATCTTACTGTATCAAAATCAA 1021
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1022 AATGGAATATGTACACATCTTCAGTATATGGTCCCTTGCATCTGTTCTAGCACAGTAC 1081
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1082 TCACCTCAGCAATTAGCAGGGAAGAATTTGAGTGTTCCTTATGGTTCGTGGTTGGCT 1141
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1142 GCCACTCTGTACTCTTTAAAGTACACACAAGATGCTACACGGGGTCTGCTTGAATAAA 1201
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1202 ATAAACAAGCAAGTTATGTATCTTAAATCAAGGCTTGATCAAGAAGTGTGGTGGCACA 1261
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle 398
Db 1262 GATGTCTTCGTGAAAAACATGAAGCTCAGAGAGACACCCATCATTTGTCACATATATT 1321
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418

Db 1322 CCCCAGGGTCAATAGATTCACTCTTTGAAGGAACGTGTAATTAGTTAGGGTGATGAA 1381
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1382 AAGCACAGAAGAACTTAGCTCGCGCTCCCACTCCAATGATGACACTTTGGATGAAGGA 1441
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysVal 458
Db 1442 GTAGGACTTGTGATTCAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAGTA 1501
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyLysHis 478
Db 1502 CCAAGACTCCCTGCCACAGCAGCAAGAACCTGAAGCAGCTGTCAATTAGTATGGGAACAT 1561
RESULT 7
US-10-236-417-191
; Sequence 191, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 191
; LENGTH: 1601
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1588)
US-10-236-417-191
Alignment Scores:
Pred. No.: 9.52e-290 Length: 1601
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 13 Gaps: 1
US-10-622-516-2 (1-478) x US-10-236-417-191 (1-1601)
QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 13 ATGCTGTGATCACTTCCTTGAATGCAGAAAGCTTGCTGGCCAAAAGATGTGGAAATTGTT 72
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 73 GCCCTTGAGATCTATTTCCTTCTCAATATGTGATCAAGACAGAGTTGGAATAATATGAT 132

QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 133 GGTGTAGATGCTGGAAAGTATACCATTTGGCTTGGCCAGGCCAAGATGGGCTTCTGCACA 192
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 193 GATAGAGAAGATATTAATCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 252
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 253 AACCTTCTCTATGATTGCATTGGCGCGCTGGAAAGTTGGAACAGAGACAATCATCGACAAA 312
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 313 TCAAAGTCTGTGAAGACTAATTGATGCACTGTTTGAAGAGTCTGGGAATACAGATATA 372
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 373 GAAGGAATCGACACAACTAATGCATGCTATGGAGGCAAGCTGCTGCTTCAATGCTGTT 432
QY 141 AsnTrpIleGluSerSerSerTyrAsp----- 149
Db 433 AACTGGATTGAGTCCAGCTCTTGGGATGGAAGGATGCCCTGGTAGTTCAGAGAGATATT 492
QY 149 ----- 149
Db 493 GCTGTATATGCCACAGGAATGCTAGACCTACAGGTGGAGTTGGAGCAGTAGCTTGCTA 552
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 553 ATTGGGCCAATGCTCCTTTAATTTTGAACGAGGCGTTCGTGGAGACATATGCAACAT 612
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 613 GCCATGATTTTACAGCCGTGATATGCTATCTGAATATCCTATAGTAGATGAAAAATC 672
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 673 TCCATACAGTGCCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACTGCCAAAAAGATC 732
QY 199 HisAlaGlnTyrGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 733 CATGCCAGTGGCAGAAAGAGGAATGATAAAGATTTAACCTTGAATGATTTTGGCTTC 792
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 793 ATGATCTTTCACTCACCATATTGTAACCTGCTTCAGAAATCTCTAGCTCGAATGTTGCTG 852
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 853 AATGACTTCCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTGGCCTGGAAGCC 912
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 913 TTTGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGAGAGGCAATTATG 972
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 973 AAGGCTAGCTCTGAACCTCTTCACTCAGAAACAAAGGCATCTTACTGTATCAAAATCAA 1032
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1033 AATGGAATATGTACACATCTTCAGTATATGGTTCCTTGCAATCTGTTAGCACAGATAC 1092
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1093 TCACCTCAGCAATTAGCAGGGAAGAAATTGGAAGTGTTCCTTATATGCTTGCTTGCGCT 1152
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1153 GCCACTCTGTACTCTCTTAAAGTCAACACAGATGCTACACCGGGGTCTGCTCTTGATAAA 1212

QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1213 ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAACTGCTGTCACCA 1272
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db 1273 GATGCTTCGCTGAAAAACATGAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT 1332
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
Db 1333 CCCAGGGTTCATATGATTCACCTCTTGAAGGAACGTGGTACTTAGTTAGGGTGATGAA 1392
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1393 AAGCAGAGAAGAACTTAAGCTCGGCGTCCCACTCCAATGATGACACTTTGGATGAAGGA 1452
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysVal 458
Db 1453 GTAGGACTTGTCATTCAAACATAGCACTGAGCATATTTCCAAGCCCTGCCAAGAAAGTA 1512
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db 1513 CCAAGACTCCCTGCCACAGCAGCAAGAACTGAAGCAGCTGTCAATTAGTAATGGGAAACAT 1572

RESULT 8
US-10-307-817-363
; Sequence 363, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Agree et al.
; TITLE OR INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307,817
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 363
; LENGTH: 1601
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1587)
US-10-307-817-363

Alignment Scores:
Pred. No.: 9.52e-290 Length: 1601
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 42
Query Match: 98.77% Indels: 1
DB: 13 Gaps: 1

US-10-622-516-2 (1-478) x US-10-307-817-363 (1-1601)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 13 ATGCCGTGATCACTTCCTTTGAATGCAGAAAGCTTGCTGGCCAAAAGATGGGAATTGTT 72
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 73 GCCCTTGAGATCTATTTCCTTCTCAATATGTGTGATCAAGCAGAGTTGGAATAATATGAT 132
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 133 GGTGTAGATGCTGGAAGATATACCATTTGGCTTGGCCAGGCCAAGATGGGCTTCTGCACA 192
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 193 GATAGAGAAGATATTAATCTCTTGCATGACTGTGTTCAGAATCTTATGGAGAGAAAT 252
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100

Db 253 AACCTTTCCTATGATTGCATTGGGGCGGCTGGAAGTTGGAACAGACAATCATCGACAAA 312
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluSerGlyAsnThrAspIle 120
Db 313 TCAAAAGTCTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGCTGGGAATACAGATATA 372
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 373 GAAGGAATGCACAACTAATGCATGCTATGAGGCACAGCTGCTGTTCAATGCTGTT 432
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 433 AACTGGATTGAGTCCAGCTCTGGGATGGACGGTATGCCCTGCTAGTTGCAGAGATATT 492
QY 149 ----- 149
Db 493 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGGAAGTTGGAGCAGTAGCTTGCTA 552
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 553 ATGGGCCAAATGCTCCTTAATTTTGAACGAGGGCTGCGGACACATATGCACACAT 612
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 613 GCCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAATACTC 672
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 673 TCCATACAGTGTACTACCTCAGTGCATTAGACCGCTGCTACTGCTACTGCAAAAACATC 732
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 733 CATGCCCAGTGGCAGAAAGGGAATGATAAAGATTTAACCTTGAATGATTTGGGCTTC 792
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 793 ATGATCTTTCACCTCACCAATTTGTAACTGGTTCAGAAATCTCTAGCTCGGATGTGCTG 852
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 853 AATGACTTCCCTTAATGACCAGAAATAGAGATAAAATAGTACTTAGTGCCCTGGAAGCC 912
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 913 TTGGGGGATGTTAATTAGAAGACACCTACTTGTATAGATGTGAGAAGGCATTATG 972
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 973 AAGGCTAGCTCTGAACCTCTTCAGTCAGAAAAACAAGGCATCTTACTTGATCAAAATCAA 1032
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1033 AATGAAATATGTACACATCTTCAGTATATGTTCCCTTCATCTGTTAGCAGACGTAC 1092
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1093 TCACCTCAGCAATTAGCAGGGAAGAAATTGGAGTGTCTTATGTTCTGTTGGCT 1152
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1153 GCCACTCTGTACTCTCTTAAGTCACACAAGATGCTACACGGGGTCTGCTCTGTATAAA 1212
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1213 ATAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGACTGCTGTGGACCA 1272
QY 379 AspValPheAlaGluAsnMetLysLeuArgGlyLysAspThrHisLysLeuValAsnTyrIle 398
Db 1273 GATGCTTCGCTGAAACATGAACTCAGAGAGACACCCATCATTTGGTCAACTATATT 1332
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
Db 1333 CCCCAGGGTTCAATAGATCACTCTTTGAAGGAACGTGTACTTAGTGGGTGATGAA 1392

QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1393 AAGCACAGAGAAGACTTACGCTCGGCGTCCCACTCCAATGATGACACTTTGATGAAGGA 1452
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db 1453 GTAGGACTTGTGCATTCAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1512
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyLysHis 478
Db 1513 CCAAGACTCCCTGCCACAGACAGAACCTGAAGCACTGTCATTAGTAATGGGAAACAT 1572

RESULT 9

US-10-236-417-181
; Sequence 181, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 181
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1582)
US-10-236-417-181

Alignment Scores:
Pred. No.: 1e-289 length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 13 Gaps: 1

US-10-622-516-2 (1-478) x US-10-236-417-181 (1-1650)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 22 ATGCCTGATCATCTTCCTTGAATGCAGAAGCTGCTGCCAAGATGTTGGGATTTGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 82 GCCCTTGAGATCATTTTCTTCTCATATATGTGATCAAGCAGAGTTGGAATAATATGAT 141

QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 142 GGTGTAGATCTGGGAAGTATACCATTTGGCTGGCCAGGCCAAGATGGGCTTCTGCACA 201
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGTTCAAGAACTTATGAGAGAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleLeuAspLys 100
Db 262 AACCTTTCCTATGATTGCATTGGGCGGCTGGAAAGTTGGAACAGACAAATCATCGACAAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGAATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 382 GAAAGGATCGACACAATAATGCATGCTATGAGGCACAGCTGCTGCTTCAATGCTGTT 441
QY 141 AsnTrpIleGluSerSerSerTyrAsp----- 149
Db 442 AACTGGATTGATGCCAGCTCTGGGATGGACGGTATGCCCTGTAGATTGCAGAGATATT 501
QY 149 ----- 149
Db 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGGAGTTGGACAGATAGCTTGCTTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATGGGGCCAAATGCTCCTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 622 GCCTATGATTTTACAGACCTGATATGCTATCTGAATATCCTATAGATGAGAGAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 682 TCCATAAGTGTACTACCTCAGTGCATAGACCGCTGCTATTCTGTCACACTGCAAAAAGATC 741
QY 199 HisAlaGlnTyrGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCAGTGGCAGAAAGAGGAATGATAAAGATTTTACCCTGAATGATTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 802 ATGATCTTTCACTCACCATATTGTAACCTGGTTCAAGAAATCTTAGCTCGGATGTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCCTTAATGACCAAGATAGAGATAAAATAGTATCTATAGTGGCCTGGAAGCC 921
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 922 TTGGGATGTTAAATTAGAAAGACACCTTACTTGATAGAGATGTGAGAAAGCCATTATG 981
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 982 AAGGCTACCTGAACTCTTCAGTCAAGAAAACAAAGGCATCTTTACTTGATCAATCA 1041
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1042 AATGGAATATGTACACATCTTCAGTATAGTTCCTTCGATCTGTTCTAGCACAGTAC 1101
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1102 TCACCTCAGCAATTAGCAGGGAAGAAATGGAAGTGTCTTTCTATGTTGGTTGGCT 1161
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1162 GCCACTCTGTACTCTTAAGTCAACACAGATGCTACACCGGGGTCTGCTCTTGATAAA 1221
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378

Db 1222 ATAAACGCAAGTTTATGTGATCTTAATCAAGGCTTGATTCAAGAACTGTGTGCACCA 1281
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle 398
Db 1282 GATGCTTCGCTGAAACATGAAAGCTCAGAGAGACACCCCATCATTTGGTCAACTATATT 1341
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
Db 1342 CCCCAGGGTTCAATAGATTCACTCTTGAAGAAACGTGTACTTAGGTGATGAA 1401
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1402 AAGCACAGAAAGACTTACGCTGGGCGTCCCACTCCAATGATGACACTTTGATGAAGGA 1461
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGlnHisIleProSerProAlaLysLysVal 458
Db 1462 GTAGGACTGTGTGATTCAAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1521
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyLysHis 478
Db 1522 CCAAGACTCCCTGCACACAGACGAACCTGAAGCAGCTGTCAATTAGTATGGGAAACAT 1581

RESULT 10
US-10-236-417-183
; Sequence 183, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 183
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1582)
US-10-236-417-183

Alignment Scores:
Pred. No.: 1e-289 Length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 13 Gaps: 1

US-10-622-516-2 (1-478) x US-10-236-417-183 (1-1650)

QY 1 MebProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 22 ATGCTGGATCATTCTTTGAAATGCAAGAAAGCTTGCGCCAAAGATGTTGGATGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 82 GCCCTTGAGATCTATTCTTCTTCAATATGTTGATCAAGCAGAGTGGAAAAATATGAT 141
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 142 GGTGTAGATGCTGGGAAGTATACCAATTGGCTTGCGCCAGGCCAAGATGGGCTTCTGCACA 201
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAACTCTTATGGAGAGAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 262 AACCTTTCTATGATTGATTCATTTGGGGCTGGAAAGTTGGAACAGACAAATCATCGACAAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAGTCTGTGAAGACTAATTGTGACAGCTGTTGAAGAGCTCGGAATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 382 GAAGGAATCGACACAACTAATGCTATGGAGGACACAGCTGCTCTTCATGCTGTT 441
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 442 AACTGGAATTGAGTCCAGCTCTTGGGATGCAAGGATGCCCTGGTAGTTGCAGAGATATT 501
QY 149 ----- 149
Db 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGAGCAGTAGCTCTGCTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATGGGCCCAATGCTCCTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 622 GCCTATGATTTTACAAAGCCTGATGCTATCTGAATATCCTATAGTAGATGGAAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 682 TCCATACAGTGTACTACCTCAGTGCATTAGACCGCTGCTATTCTCTACTGCCAAAAAGATC 741
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCCAGTGGCAGAAAGAGGAAATGATAAAGATTTACCTTGAATGATTTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 802 ATGATCTTTCACTCACCATATTGTAACTGCTTCAGAAATCTCTAGCTCGATGTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCCTTAATGACACAGAAATAGATATAAATAAGTATCTAATAGTGGCTGGAAGCC 921
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 922 TTGGGGATGTTAAATTAGAAAGACCTACTTTGATAGAGATGGAGAAGGCAATTTATG 981
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 982 AAGGCTAGCTCTGAACCTTTCAGTCAAGAAAACAAAGGCATCTTTACTTGTATCAAAATCA 1041
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1042 AATGGAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTCTTAGCACAGATAC 1101

QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1102 TCACCTCAGCAATTAGCAGGGAAGAATTTGAGTGTCTTATGTTCTGTGTTGGCT 1161
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1162 GCCACTGTACTCTCTTAAAGTCAACAAGATGCTACACCGGGGCTGCTCTTGATATA 1221
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1222 ATAACAGCAAGTTATGTATCTTAAATCAAGGCTTGATTCAAAGAACTGGTGGCACA 1281
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db 1282 GATGCTTCGCTGAAAAACATGAAGCTCAGAGAGACACCATCATTTGGTCAACTATATT 1341
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
Db 1342 CCCAGGGTTCATATAGATTCACCTTTGAAGGAACGTGGTACTTAGTGGTGATGAA 1401
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspThrLeuAspGluGly 438
Db 1402 AAGCACAAGAAGACTTACCGTCCGGCTCCCACTCCAAATGATGACACTTTGGATGAAGA 1461
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysIleVal 458
Db 1462 GTAGGACTGTGCATTCAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1521
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db 1522 CCAAGACTCCCTGCCACAGCAGCAAACTGAAGCAGCTGCATTAGTAATGGGAACAT 1581

RESULT 11

US-10-236-417-185
: Sequence 185, Application US/10236417
: Publication No. US20040048256A1
: GENERAL INFORMATION:
: APPLICANT: Agee et al.
: TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
: FILE REFERENCE: 21402-442C
: CURRENT APPLICATION NUMBER: US/10/236,417
: PRIOR FILING DATE: 2003-01-06
: PRIOR APPLICATION NUMBER: US60/318,120
: PRIOR FILING DATE: 2001-09-01
: PRIOR APPLICATION NUMBER: US60/318,430
: PRIOR FILING DATE: 2001-09-10
: PRIOR APPLICATION NUMBER: US60/322,781
: PRIOR FILING DATE: 2001-09-17
: PRIOR APPLICATION NUMBER: US60/318,184
: PRIOR FILING DATE: 2001-09-07
: PRIOR APPLICATION NUMBER: US60/361,663
: PRIOR FILING DATE: 2002-03-05
: PRIOR APPLICATION NUMBER: US60/396,412
: PRIOR FILING DATE: 2002-07-17
: PRIOR APPLICATION NUMBER: US60/322,636
: PRIOR FILING DATE: 2001-09-17
: PRIOR APPLICATION NUMBER: US60/322,817
: PRIOR FILING DATE: 2001-09-17
: PRIOR APPLICATION NUMBER: US60/322,816
: PRIOR FILING DATE: 2001-09-17
: PRIOR APPLICATION NUMBER: US60/323,519
: PRIOR FILING DATE: 2001-09-19
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 341
: SOFTWARE: Custom
: SEQ ID NO 185
: LENGTH: 1650
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (22) .. (1582)

US-10-236-417-185

Alignment Scores:
Pred. No.: 1e-289 Length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 13 Gaps: 1

US-10-622-516-2 (1-478) x US-10-236-417-185 (1-1650)

QY 1 MetProGlySerLeuProLeuAenAlaGluAlaCysTrpProLysAspValGlyIleVal 20
|||
Db 22 ATGCGTGGATCACTTCCCTTGAAATGCAAGAGCTTGCCAAAGAATGTTGGGATTGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
|||
Db 82 GCCCTTGAGATCTATTCTCTTCAATATGTTGATCAAGCAGAGTTGAAAAATATGAT 141
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
|||
Db 142 GGTGTAGATGCTGGGAAGTATACCATTTGGCTTGGCCAGGCCCAAGATGGGCTTTCGCACA 201
QY 61 AspArgGluAspIleasnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
|||
Db 202 GATAGAGAAGATATTAACTCTCTTTCATGACTGTGGTTCAGAAATCTTAATGAGAGAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
|||
Db 262 AACCTTTCCTATGATTGCAATTTGGCGGCTGGAAGTTGGAACAGACAAATCATGCACAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
|||
Db 322 TCAAGCTCTGTGAAGACTAATTGATGCAAGCTGTTGAAGAAGCTCGGAATATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
|||
Db 382 GAAGGAATCGACACAACTAATGCATGCTATGAGGACAGAGCTGCTGTCTCAATGCTGTT 441
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
|||
Db 442 AACTGATTGAGTCCAGCTCTTGGAGTGAACGGTATGCCCTGTAGTTGACAGAGATATT 501
QY 149 ----- 149
Db 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGCAGTACCTCTGCTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
|||
Db 562 ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGACACATATGCAACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
|||
Db 622 GCCTATGATTTTACAAAGCCTGATATGCTATCTGAATATCTATAGTAGTGGGAAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
|||
Db 682 TCCATACAGTGTCTACCTCAGTGCATTAGACCGCTGCTATTCCTACTGCAAAAAGATC 741
QY 199 HisAlaGlnTrpGlnLysGluLysAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
|||
Db 742 CATGCCCAAGTGCAGAAAGGGAATGATAAAGATTTTACCTTGAATGATTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
|||
Db 802 ATGATCTTTCACCTACCATATTTGTAACCTGTTCAGAAATCTCTAGCTCGGATGTTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
|||
Db 862 AATGACTTCTTAATGACCAAGATAGAGATAAATAAGTATCTATAGTGCGCTGGAAAGCC 921
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278

Db 922 TTTGGGAGTCTTAATTAGAGACACCTTCTTGTAGAGATGTGAGAGGCACTTATG 981
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
|||
Db 982 AAGGCTAGCTCTGAACCTCTTCAAGTCAAGAAACAAAGGCATCTTACTTGTATCAAAATCAA 1041
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
|||
Db 1042 AATGGAATATGTACACATCTTCAATATATGTTCCCTTGCAATCTGTTCTAGCACAGTAC 1101
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
|||
Db 1102 TCACCTCAGCAATTAGCAGGGAAGAAATGAGATGTTTCTTAATGCTTCTGTTGGCT 1161
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
|||
Db 1162 GCCACTCTGTACTCTCTTAAAGTCAACACAGATGCTACACCGGGCTGCTCTTGAATAA 1221
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
|||
Db 1222 ATACACGCAAGTTATATGTATCTTAATCAAGGCTTGATTCAAGAACTGCTGCGCACA 1281
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisLeuValAsnTyrIle 398
|||
Db 1282 GATGCTTCGCTGAATAACATGAAAGCTCAGAGAGACACCATCATTTGGTCAACTATATT 1341
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
|||
Db 1342 CCCCAGGGTCAATAGATTCACTCTTGAAGAACTGTACTTACTTAAAGGTGATGAA 1401
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
|||
Db 1402 AAGCACAAGAAAGAACTTACGCTCGGGCTCCACTCCAATGATGACACTTTGGATGAAGGA 1461
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
|||
Db 1462 GTAGGACTTGTGCAATCAAAATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1521
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyLysHis 478
|||
Db 1522 CCAAGACTCCCTGCCACAGCAGCAAGAACCTGAAGCAGCTGTCAATTAATGGGAACAT 1581
RESULT 12
US-10-236-417-189
; Sequence 189, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19

; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 193
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22) .. (1582)
US-10-236-417-193

Alignment Scores:
Pred. No.: 1e-289 Length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 13 Gaps: 1

US-10-622-516-2 (1-478) x US-10-236-417-193 (1-1650)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 22 ATGCTTGATCATTCTTCTTGAATGCAGAGCTTCTGCCCAAGAAGATGTGGATGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 82 GCCCTTGAGATCTATTCTTCTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 141
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 142 GGTGTAGATGCTGGAGATATACCATTTGGCTTGGCCAGGCCAAGATGGCTTCTGCACA 201
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATTAACTCTTGTGCACTGTGTCAGAAATCTTATGAGAGAAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 262 AACCTTCTCATGATTGCAATGGCGGGCTGGAAGTTGGAACAGACAATCATCGACAAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAGAGTCTGTGAAGACTTAATTGATGCACTGTTGAAAGAGTCTGGGAATACAGATATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 382 GAAAGAATGCACAACTAATGCAATGCTATGAGGACACAGCTGCTGTTCAATGCTGTT 441
QY 141 AsnTyrIleGluSerSerSerTyrAsp----- 149
Db 442 AACTGATTAAGTCCAGCTTTGGGATGACGGTATGCCCTGTAGTTGACGAGATATT 501
QY 149 ----- 149
Db 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGACAGTACTGTGCTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATTGGGCAAAATGCTCTTAATTTTGAACGAGGGCTTGTGGGAACACATATGCAACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178

Db 622 GCCTATGATTTTAAAGCCCTGATATGCTATCTGAATATCCTATAGTAGATGGGAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIle 198
Db 682 TCATACAGTGTACTTCACTCAGTGCATTAGACCCGCTGCTATTCTGTCTACTGCAAAAAGATC 741
QY 199 HisAlaGlnTyrGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCCACTGGCAGAAAGAGGAAATGATAAGATTTTACCCTTGAATGATTTGGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeu 238
Db 802 ATGATCTTTCACCTCACCATATTTGTAAGTGTTCAGAAATCTCTAGCTCGGATGTTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCTTAATGACCAGAAATAGAGATAAAATAGTATCTATAGTGGCTGGAAGCC 921
QY 259 PheGlyAspValLysLeuGlyAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 922 TTTGGGATGTTAAATTAGAAAGACCTACTTGTAGAGATGTGGAGAAAGCATTATATG 981
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 982 AAGGCTAGCTCTGAACCTCTTCAGTACAGAAACAAAGGCATCTTACTTGTATCAAAATCAA 1041
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1042 AATGAAATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTCTACACACAGTAC 1101
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1102 TCACCTCAGCAATTAGCAGGAGAGAAATTGAGTGTCTTATGTTCTGTTGGCT 1161
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1162 GCCACTGTACTCTCTTAAAGTCACACAAGATGCTACACGGGCTGTGCTCTTGATATAA 1221
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1222 ATAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATCAAGAACTGTGTGGACCA 1281
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisLysLeuValAsnTyrIle 398
Db 1282 GATGTCTTCGCTGAAAACATGAAGCTCAGAGAGACACCATCATTTGGTCAACTATATT 1341
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrPyrLeuValArgValAspGlu 418
Db 1342 CCCAGGGTTCAATAGATTCACTCTTGAAGGAACGTGTACTTAGTTAGGTGATGAA 1401
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1402 AAGCACAGAAGAACCTTACCGCTCGGCTCCCACTCCAATGATGACACTTTGATGAAGGA 1461
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db 1462 GTAGGACTTGTGCATTCAAACATAGCACTGAGCATATTCAGCCCTGCCAAGAAAGTA 1521
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyLysHis 478
Db 1522 CCAAGACTCCTGCCACAGCAGACGAACCTGAAGCAGCTGTCAATTAGTAATGGGAAACAT 1581

RESULT 14
US-10-236-417-197
; Sequence 197, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06

; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 197
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1582)
US-10-236-417-197

Alignment Scores:
Pred. No.: 1e-289 Length: 1650
Score: 2480.00 Matches: 478
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 0
Query Match: 98.77% Indels: 42
DB: 13 Gaps: 1

US-10-622-516-2 (1-478) x US-10-236-417-197 (1-1650)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 22 ATGCCCTGGATCACTTCCTTGAATGCAGAGCTTGCTGGCCAAAGATGTTGGGATTGTT 81
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 82 GCCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAATATGAT 141
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 142 GGTGTAGATGCTGGAGATATACCATTTGGCTTGGCCAGGCCAAGATGGCTTCTGCACA 201
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATTAACTCTCTTTCATGACTGCTGTTCAAGATCTTATGAGAGAAT 261
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 262 AACCTTTCCTATGATGTCATGGCGGCTGGAAGTTGGAACAGACAAATCATCGACAAA 321
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAACTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATA 381
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 382 GAAGGAATCGACACAACTAATGCTATGAGGACACAGCTGCTGCTTCAATGCTGTT 441
QY 141 AsnTrpIleGluSerSerSerTyrAsp----- 149
Db 442 AACTGATTGAGTCAGCTCTTGGGATGCAGCGTATGCCCTGTAGTTGCAGAGATATT 501

QY 149 ----- 149
Db 502 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGGAGCAGTAGCTTGTCTA 561
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATTGGGCCAAATGCTCTTTAATTTTGAACGAGGGCTTCGTGGACACATATGCAACAT 621
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 622 GCCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGGAACTC 681
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIle 198
Db 682 TCCATACAGTGTACCTCAGTGCATTAGACCGCTGCTATCTGTACTGCATAAAGATC 741
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCCAGTGGCAGAAAGAGGAAATGATTAAGATTTTACCTTGAATGATTTTGCTTC 801
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 802 ATGATCTTTCACCTCACCAATATTGTAACCTGGTTCAGAATCTTAGCTCGGATGTTGCTG 861
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCCTTAATGACCAGAAATAGAGATAAAATAGTATAGTGCGCTGGAAGCC 921
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 922 TTTGGGATGTTAAATTAGAAGACACCTACTTGTATAGATGTGAGAGGCAATTATG 981
QY 279 LysAlaSerSerGlnLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 982 AAGGCTAGCTCTGAACCTCTTCAAGTCAAGAAACAAAGGCATCTTACTGTATCAATCAA 1041
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1042 AATGGAATATGTACACATCTTCAATATAGTTCCTTGCATCTGTTCTAGCACAGTAC 1101
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1102 TCACCTCAGCAATTAGCAGGAAAGAAATGAGTGTCTTATAGTTCTGTTGGCT 1161
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1162 GCCACTCTGTACTCTCTTAAAGTCACACAAAGATGCTACACGGGCTGCTCTGTGATAAA 1221
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1222 ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGCTGTGTGCACCA 1281
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db 1282 GATGCTTCGCTGAAAACATGAAGCTCAGAGAGGACCCCATTCATTGTGTCACTATATT 1341
QY 399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrLeuValArgValAspGlu 418
Db 1342 CCCCAAGGTTCAATAGATTCACCTTTGAAGGAACGTGTACTTAGTGGGTGATGAA 1401
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 1402 AAGCACAGAAAGAACTTACGCTCGCGCTCCCACTCCAATGATGACACTTTGATGAAGA 1461
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysVal 458
Db 1462 GTAAGACTTGTGATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1521
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db 1522 CCAAGACTCCCTGCCACGACGACAGAACCTGAAGCAGCTGTCAATTAGTATGGGGAACAT 1581

Db	1402	AAGCAGAGAACTTACGCTCGGCGTCCCACTCCAATGATGACACTTGGATGAAGA	1461
QY	439	ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysIysVal	458
Db	1462	GTAAGACTTGTGCATTCAACATAAGCAACTGAGCATATATCCAAGCCCTGCCAAGAAAGTA	1521
QY	459	ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis	478
Db	1522	CCAAGACTCCCTGCCACAGCAGAGAACCCTGAAGCAGCTGTTCATTAGTAATGGGGAACAT	1581

Search completed: June 24, 2004, 17:13:40
Job time : 577 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 24, 2004, 13:29:18 ; Search time 115 Seconds
(without alignments)
2306.665 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPEGLPLNAEACWPKDVGIV.....PRLPATAAEPEAAVISNGEH 478

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_P/US10622516/runat_23062004_162634_673/app_query.fasta_1.647
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=DCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10622516_QCGN_1_1_105 @runat_23062004_162634_673 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	2511	100.0	2002	4	US-09-819-993-1 Sequence 1, Appli
2	2511	100.0	2002	4	US-10-193-295-1 Sequence 1, Appli
3	2089	83.2	1824	1	US-08-305-505-1 Sequence 1, Appli
4	839.5	33.4	28001	4	US-09-819-993-3 Sequence 3, Appli
5	839.5	33.4	28001	4	US-10-193-295-3 Sequence 3, Appli
6	771	30.7	615	4	US-09-370-838-245 Sequence 245, App
7	703	28.0	500	4	US-09-370-838-128 Sequence 128, App
8	686	27.3	506	4	US-09-401-064-187 Sequence 187, App
9	585	23.3	4775	3	US-09-306-595C-1 Sequence 1, Appli
10	585	23.3	4775	4	US-09-925-388-1 Sequence 1, Appli
11	530.5	21.1	635	3	US-08-998-416-730 Sequence 730, App
12	395	15.7	472	4	US-09-833-381-1648 Sequence 1648, App

13	315	12.5	1167	4	US-09-107-532A-1715	Sequence 1715, Ap
14	301.5	12.0	10813	4	US-08-956-171E-364	Sequence 364, App
15	284.5	11.3	1311	4	US-09-134-001C-1458	Sequence 1458, Ap
16	282	11.2	1350	4	US-09-134-000C-2039	Sequence 2039, Ap
17	268	10.7	305	4	US-09-313-294A-4313	Sequence 4313, Ap
18	262.5	10.5	448	4	US-09-833-381-1645	Sequence 1645, Ap
19	255.5	10.2	288	4	US-09-313-294A-5701	Sequence 5701, Ap
20	244	9.7	18436	4	US-08-961-527-87	Sequence 87, Appl
21	242	9.6	307	4	US-09-833-381-744	Sequence 744, Appl
22	209.5	8.3	951	4	US-09-241-750-1	Sequence 1, Appli
23	198	7.9	385	4	US-09-833-381-1646	Sequence 1646, Ap
24	178.5	7.1	293	4	US-09-313-294A-7134	Sequence 7134, Ap
25	140.5	5.6	263	4	US-09-313-294A-686	Sequence 686, App
26	136	5.4	232	4	US-09-833-381-745	Sequence 745, App
27	134.5	5.4	421	4	US-09-833-381-1647	Sequence 1647, Ap
28	123.5	4.9	4488	4	US-09-543-681A-1814	Sequence 1814, Ap
29	123	4.9	1077	4	US-09-543-681A-3962	Sequence 3962, Ap
30	122	4.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
31	105.5	4.2	3756	4	US-09-813-742A-2	Sequence 2, Appli
32	105.5	4.2	3760	1	US-08-261-663A-3	Sequence 3, Appli
33	105.5	4.2	3760	4	US-09-357-206A-2	Sequence 2, Appli
34	105.5	4.2	3760	5	PCT-US95-07754A-3	Sequence 3, Appli
35	104	4.1	16592	4	US-08-956-171E-53	Sequence 53, Appl
36	103.5	4.1	9510	4	US-09-453-702B-256	Sequence 256, App
37	103	4.1	400	4	US-08-956-171E-4047	Sequence 4047, Ap
38	102.5	4.1	2412	1	US-08-158-232-9	Sequence 9, Appli
39	102.5	4.1	2412	1	US-08-304-626-9	Sequence 9, Appli
40	102.5	4.1	2412	1	US-08-316-301A-11	Sequence 11, Appl
41	102.5	4.1	2412	1	US-08-611-928-9	Sequence 9, Appli
42	102.5	4.1	2412	3	US-09-173-891-9	Sequence 9, Appli
43	102.5	4.1	2412	3	US-09-076-137-11	Sequence 11, Appl
44	102.5	4.1	2412	4	US-09-738-363-11	Sequence 11, Appl
45	102.5	4.1	2412	5	PCT-US92-03624-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-819-993-1
; Sequence 1, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819, 993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Human
US-09-819-993-1

Alignment Scores:

Pred. No.: 2.6e-302 Length: 2002
Score: 2511.00 Matches: 478
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-622-516-2 (1-478) x US-09-819-993-1 (1-2002)

Qy 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProlyAspValGlyIleVal 20
Db 145 ATGCCTGGATCACTTCCTTGAATGCAAGAAGCTTGTGCCCAAAAGATGTGGAAATTGTT 204
Qy 21 AlaleuGluIleTyRPhenProSerGlnTyRValAspGlnAlaGluLeuGluTyRAsp 40

Db	205	GCCCTTGAGATCTATTTCTCTCTCATATATGTTGATCAAGCAGAGCTTGAAAAATATGAT	264
QY	41	GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
Db	265	GGTGTAGATGCTGGAAAGATATACCATTTGGCTTGGCCAGGCCAAGATGGGCTTCTGCACA	324
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn	80
Db	325	GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGCTTCAGAATCTTATGGAGAGAAAT	384
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
Db	385	AACCTTTCCTATGATTGCATTGGCGCGCTGGAACTTGGAAACAGACAAATCATCGACAAA	444
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
Db	445	TCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGAATACAGATATA	504
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
Db	505	GAAAGAAATCGACACAACATAATGCTATGAGGACACAGCTGCTGTCTTCATGCTGTT	564
QY	141	AsnTyrPileGluSerSerSerTyrAspGlyLeuArgGlyThrHisMetGlnHisAlaTyr	160
Db	565	AACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGTGGACACATATGCACAACATGCCTAT	624
QY	161	AspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIle	180
Db	625	GATTTTACAGCCTGATATGCTATCTGAATATCTTATAGATGGAACCTCTCCATA	684
QY	181	GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAla	200
Db	685	CAGTGTACTCTCAGTGCATTAGACCGCTGCTACTCTGTCTACTGCAAAAAGATCCATGCC	744
QY	201	GlnTyrGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIle	220
Db	745	CAGTGGCAGAAAGAGGAAATGATAAGATTTTACCTTGAATGATTTTGGCTCATGATC	804
QY	221	PheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp	240
Db	805	TTTCACTCACCATATTGTAACCTGGTTCAGAAATCTCTAGCTGGATGTTGCTGAATGAC	864
QY	241	PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGly	260
Db	865	TTCTTAATGACCAAGATAGATATAAAATAGTATCTATAGTGGCCTGGAAGCCTTTGGG	924
QY	261	AspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAla	280
Db	925	GATGTTAAATTAGAAAGACACCTTACTTGTATAGATGTGAGAAAGCATTTTGAAGGCT	984
QY	281	SerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGly	300
Db	985	AGCTCTGAACCTCTTCAGTCAGAAAAACAAAGGCATCTTACTGTATCAAAATCAAAATGGA	1044
QY	301	AsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerPro	320
Db	1045	AATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTGTTCTAGCACAGTACTCACT	1104
QY	321	GlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThr	340
Db	1105	CAGCAATTAGCAGGGAAGAGATTGAGTGTCTTATGTTCTGTTGGCTGCCACT	1164
QY	341	LeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThr	360
Db	1165	CTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGCTGCTCTTGATAAATAACA	1224
QY	361	AlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAspVal	380
Db	1225	GCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAGAAGACTGTTGACACAGATGTC	1284
QY	381	PheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIleProGln	400
Db	1285	TTCGCTGAAAAATGAAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATFTCCCCAG	1344

QY	401	GlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGluLysHis	420
Db	1345	GGTTCATAGATTCACTCTTTGAAGGAACGTGTACTTAGTGGGTGATGAAAAACAC	1404
QY	421	ArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGlyValGly	440
Db	1405	AGAAGAACTTACGCTCGCGCTCCCACTCCAAATGATGACACTTTGGATGAAGAGTAGGA	1464
QY	441	LeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysValProArg	460
Db	1465	CTTGTCATTCAAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTACCAAGA	1524
QY	461	LeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyLysHis	478
Db	1525	CTCCCTGCCACAGCAGACAGAACTGAAGCAGCTGTCTATTAGTATGGGAAACAT	1578
RESULT 2			
US-10-193-295-1			
; Sequence 1, Application US/10193295			
; Patent No. 6620608			
; GENERAL INFORMATION:			
; APPLICANT: GONG, Fangcheng et al.			
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC			
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES			
; TITLE OF INVENTION: THEREOF			
; FILE REFERENCE: CL001195DIV			
; CURRENT APPLICATION NUMBER: US/10/193,295			
; CURRENT FILING DATE: 2002-07-12			
; PRIOR APPLICATION NUMBER: 08/819,993			
; PRIOR FILING DATE: 2001-03-29			
; NUMBER OF SEQ ID NOS: 5			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 1			
; LENGTH: 2002			
; TYPE: DNA			
; ORGANISM: Human			
US-10-193-295-1			
Alignment Scores:			
Pred. No.: 2.6e-302 Length: 2002			
Score: 2511.00 Matches: 478			
Percent Similarity: 100.00% Conservative: 0			
Best Local Similarity: 100.00% Mismatches: 0			
Query Match: 100.00% Indels: 0			
DB: 4 Gaps: 0			
US-10-622-516-2 (1-478) x US-10-193-295-1 (1-2002)			
QY	1	MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal	20
Db	145	ATGCCCTGATCACTTCTTGAATGCAGAGCTTGCTGCCAAAGATGTGGGAATTGTT	204
QY	21	AlaLeuGlnIleTyrPheProSerGlnTyrValAspGlnAlaGlnLeuGlnLysTyrAsp	40
Db	205	GCCCTTGAGATCTATTTCTTCTCAATATGTTGATCAAGCAGAGTTGGAATAATATGAT	264
QY	41	GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60
Db	265	GGTGTAGATGCTGAAAGTATTAACATTTGGCTTGGCCAGGCCAAGATGGGCTTGCACA	324
QY	61	AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn	80
Db	325	GATAGAGAAGATATTAACTCTTTTGCATGACTGTGTTCAGAATCTTATGGAGAGAAAT	384
QY	81	AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100
Db	385	AACCTTTCCTATGATTGCATTGGCGCGCTGGAACTTGGAAACAGACAAATCATCGACAAA	444
QY	101	SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120
Db	445	TCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA	504

QY 121 GlnGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 505 GAAGGAATCGACACAACATAATGATGCTATGAGGACAGCTGCTGCTTCAATGCTGTT 564
QY 141 AsnTrpIleGluSerSerSerTrpAspGlyLeuArgGlyThrHisMetGlnHisAlaTyr 160
Db 565 AACTGATTGAGTCCAGCTCTGGGATGGGCTTCGTGGACACATATGCAACATGCTAT 624
QY 161 AspPheTyrLysProAspMetLeuSerGlyTyrProIleValAspGlyLysLeuSerIle 180
Db 625 GATTTTACAGACCTGATATGCTATCTGAATATCCTATAGATGGAACACTCTCCATA 684
QY 181 GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAla 200
Db 685 CAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACTGCAAAAAGATCCATGCC 744
QY 201 GlnTrpGlnLysGlnGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIle 220
Db 745 CAGTGGCAGAAAGAGGAAATGATAAAGATTTTACCCTTGAATGATTTGGCTTCATGATC 804
QY 221 PheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp 240
Db 805 TTTCACCTCACCATATTTGTAACCTGTTCAGAAATCTTAGCTCGGATGTTGCTGAATGAC 864
QY 241 PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGlnAlaPheGly 260
Db 865 TTCCTTAATGACCAGATAGATAAAATAGTATCTATAGTGGCCTGGAAGCCTTGGG 924
QY 261 AspValLysLeuGlnAspThrTyrPheAspArgAspValGlnLysAlaPheMetLysAla 280
Db 925 GATGTTAAATTAGAAAGACACTTACTTTGATAGAGATGTGAGAAAGCATTTATGAAGCT 984
QY 281 SerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGly 300
Db 985 AGCTCTGAACCTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTATCAAAATCAAAATGGA 1044
QY 301 AsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerPro 320
Db 1045 AATATGTACACATCTTCAGTATATGTTCCCTTGCACTGTTCTAGCACAGTACTGACCT 1104
QY 321 GlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThr 340
Db 1105 CAGCAATTAGCAGGAGAGAAATTGAGTGTCTTCTATGTTCTGTTGGCTGCCACT 1164
QY 341 LeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThr 360
Db 1165 CTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGCTGCTCTTGATAAATTAACA 1224
QY 361 AlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAspVal 380
Db 1225 GCAAGTTTATGTATCTTAAATCAAGGCTTGATTCAAGAACTGCTGTGGCACACAGATGTC 1284
QY 381 PheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIleProGln 400
Db 1285 TTGCTGAAAAACATGAAGCTCAGAGAGACACCCCATCATTTGTCAACTATATCCCCAG 1344
QY 401 GlySerIleAspSerLeuPheGlyGlyThrTyrTyrLeuValArgValAspGlyLysHis 420
Db 1345 GGTTCATATAGATTCACCTTTGAAGAAAGCTGTACTTAGTGAAGGTGATGAAAAGCAC 1404
QY 421 ArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGlyLysValGly 440
Db 1405 AGAAGAACTTACGCTCGCGCTCCCACTCCAATGATGACACTTTGGATGAAGAGATGAGA 1464
QY 441 LeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysValProArg 460
Db 1465 CTTGTGCATTCAACATAGCAACTGAGCATATTTCCAAGCCCTGCCAAGAAAGTACCAAGA 1524
QY 461 LeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db 1525 CTCCTGCCACAGCAGCAGAACTGGAAGCAGCTGTCAATTAGTAATGGGAAACAT 1578
RESULT 3

US-08-305-505-1
; Sequence 1, Application US/08305505
; Patent No. 5668001
; GENERAL INFORMATION:
; APPLICANT: Miziorko, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGUTARYL-CoA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
; TITLE OF INVENTION: STABILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,040
; FILING DATE: 02 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 65-053-9083-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-305-505-1

Alignment Scores:
Pred. No.: 8.76e-250 length: 1824
Score: 2089.00 Matches: 399
Percent Similarity: 83.33% Conservative: 36
Best Local Similarity: 76.44% Mismatches: 43
Query Match: 83.19% Indels: 44
DB: 1 Gaps: 3

US-10-622-516-2 (1-478) x US-08-305-505-1 (1-1824)
QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 2 ATGGCTGGGTCTCTTCAGTGAACACTGAATCCTGCTGCCCCAAAGATGTGGGTATTGTT 61
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGlnLysTyrAsp 40
Db 62 GCACGTGAATCTATTTCCTCTCAGTATGTGCAGCAGACTGAGCTGAGAGAATGAC 121
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 122 GGTGTGATGCAGGCAATATACCACTTGGGTTAGGCCAGTCAAAAGATGGGCTTCTGCTCT 181
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 182 GACCGAGAGATATCAATTCCTCTGTTTGACTGTTCAGAACTTATGAGAGGAAAC 241
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 242 AGCCTTTCCTATGACTGCTATGGGAGACTGGAAGTTGGAACGAGACAATTAATTGATAA 301

QY	101	SerIysSerValIysThrAsnLeuMetGlnLeuPheGluSerGlyAsnThrAspIle	120
Db	302	TCAAAATCGGTGAAGACTGTCCTGATGCAGCTATTGGAAGATCTGGTAATACAGATGTA	361
QY	121	GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal	140
Db	362	GAAGGAATTGACACAAACCAATGCGTCTATGAGAGCACTGCTCTTTTAAATGCTATT	421
QY	141	AsnTrpIleGluSerSerSerTrpAsp-----	149
Db	422	AACGGATTGAGTCCAGTCTTGGGATGGACGCTATGCACCTGTGTCTGTGAGACATT	481
QY	149	-----	149
Db	482	GCTGTGTATGCCACTGGAAATGCGCAGGCCAACAGGTTGAGCTGTGCTGTATGCTA	541
QY	150	-----GlyLeuArgGlyThrHisMetGlnHis	158
Db	542	GTTGGGTCAAAATGCTCTTTAATTTTGGAGAGAGATTGCGTGGAAACCATGCATGCAGCAT	601
QY	159	AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu	178
Db	602	GCTTATGACTTCTATTAACACAGATATGGTTTCTGAATATCCTGTAGTTGATGGCAAAC	661
QY	179	SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle	198
Db	662	TCTATACAGTGTCTTACCTCAGTGCATTAGACCGCTGCTATAGTGTATTATCGCAATTAATC	721
QY	199	HisAlaGlnTyrGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe	218
Db	722	CATGCCCCAGTGGCAAAAAGAGGGGACAGACAGAGSTTTCACCTTGAATGATTTTGGATT	781
QY	219	MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu	238
Db	782	ATGATCTTTCAATCTCCCTACTGTAACCTGTACAGAAAGTCGGTGGCAAGACTGTGCTG	841
QY	239	AsnAspPheLeuAsnAspGlnAsnArgAsp--LysAsnSerIleTyrSerGlyLeuGlu	257
Db	842	AATGACTTTCTCAGTGACCCAGAAATGCAGAAACAGCAAAATGATGTTTCACTGTCGAA	901
QY	258	AlaPheGlyAspValLysLeuGlnAspThrTyrPheAspArgAspValGluLysAlaPhe	277
Db	902	GCTTTCAGGAGATGAAGCTTGAAGATACATATTTGATAGGGAATGTGAAAAAGCTTTT	961
QY	278	MetLysAlaSerSerGlnLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsn	297
Db	962	ATGAAAGCTAGTGCAGAGCTCTTCAATCAGAAAAACAAAGCTTCTTACTGTGTCCAAT	1021
QY	298	GlnAsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGln	317
Db	1022	CAGAATGGAACATGTACACGCCCTTCACTACGCTTGCTTCTCTTCTTAAGCCCAAG	1081
QY	318	TyrSerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeu	337
Db	1082	TACTCTCCAGACCACTTGACAGACAAAGAATCAGTGTATCATATGGCTGTGTTTT	1141
QY	338	AlaAlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAsp	357
Db	1142	GCTGCTACGCTGTATTCCATCAGAGTTTACACAGGATGCACCTCGTGTCTGCGCTTGAC	1201
QY	358	LysIleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAla	377
Db	1202	AAAATAACTGCTAGCCTTCTGTGATCTTAAAGCAAGACTTGACTACGAAATGCAATTGCA	1261
QY	378	ProAspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleValAsnTyr	397
Db	1262	CCTGATGCTTGTCTGAAAAACATGAAGATTAGACAGAGACACATCACTTGCCAACTAT	1321
QY	398	IleProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAsp	417
Db	1322	ATTCACACAGTGTTCAGTAGAAGATCTTTTGAGGGAACATGTTATCTTGTGCGTGTGAT	1381

[illegible]

```

RESULT 4
US-09-819-993-3
; Sequence 3, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 28001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(28001)
; OTHER INFORMATION: n = A,T,C or G
US-09-819-993-3

```

Alignment Scores:	
Pred. No.:	5.76e-92
Score:	839.50
Percent Similarity:	19.23%
Best Local Similarity:	19.16%
Query Match:	33.43%
DB:	4
Length:	28001
Matches:	302
Conservative:	1
Mismatches:	3
Indels:	1271
Gaps:	6

US-10-622-516-2 (1-478) x US-09-819-993-3 (1-28001)

QY	150	GlyLeuArgGlyThrHisMetGlnHisAlaTyrAspPheTyrLysProaspMetLeuSer	169
Db	18664	GGGCTTCGTGGACACATATGCAACATGCCTATGATTTTACAAGCCGTGATATGCTATCT	18723
QY	170	GluTyrProIleValAspGlyLysLeuSerIleGlnCysTyrLeuSerAlaLeuAspArg	189
Db	18724	GAATATCCTATAGTAGATGGAACCTCCCATACAGTGCTACCTCAGTCATTAGACCGC	18783
QY	190	CysTyrSerValTyrCysLysLysIleHisAlaGlnTrpGln-Lys-----	204
Db	18784	TGCTATTCTGTCTACTGCAGAAAAGATCCATGCCCCAGTGGCAGAAAGGTAAGTTTACC	18843
QY	204	-----	204
Db	18844	TTTTCCCTGGTTTGGTATGAGTTGAGAGCAGCTTAATGTACTAGGTATCTTGGTAGGC	18903
QY	204	-----	204
Db	18904	AACTACTTTGGGGCATTCTTCATTTAATATCCTTTTACCATTAATTCCTCATTCACCAA	18963
QY	204	-----	204
Db	18964	ACAACATTTCCCATAGTTCTTCGGAAAAGTGTAATTTACTAGAGAGGTAACCTTTGGAA	19023

QY 204 ----- 204
Db 19024 CTGAGGTGATCTTCGAAAAATATTAGTCCGTTTACCCTTGAAGAAATCAAGT 19083
QY 204 ----- 204
Db 19084 GGAGAAAAAGAGTGAATTGTTGTCATCTTTGAGAGAGGATTTTACAGAGTT 19143
QY 204 ----- 204
Db 19144 TTGGA CTACAGCTGTGATTCAGGAAAGCTAATGAAAAATGAATTACTAAAGTCACTTAC 19203
QY 204 ----- 204
Db 19204 CCCAAAAATATCTTTTGCACCTTGACCTGTGAATTTGTATTTGTTTTTACTGTATC 19263
QY 204 ----- 204
Db 19264 ATTAATCTGGAATTTGTGAGGCACTGAAGACAGTATTTGAGTTAATGCTATCAATA 19323
QY 204 ----- 204
Db 19324 CACATTTATACATAAAGTAACTTTTCTGTAGTCCAACTTTGCTTTTAGAGGTTATGA 19383
QY 204 ----- 204
Db 19384 GAAGGGGTTAAAAATCATATTCATATGACAATATCAGTGAATTTAGTCGCTCTGATAAG 19443
QY 204 ----- 204
Db 19444 AAGCATTTCTGCAGTATATATTAACAGAATAGTGTTTCTAACTTTTATTAGSACCC 19503
QY 204 ----- 204
Db 19504 ACAGTAAGAAGTACATGTTAATTGTATGTATGCCAGACTGAACAACAAAATGTCATGA 19563
QY 204 ----- 204
Db 19564 CATTACTTACCCTTGCTGCAAGTTATTCAGTTGCTATTTTCTACTGCATTTGTTTT 19623
QY 204 ----- 204
Db 19624 TAAAAATACTCTTTTATTTAAAAAAATACTAATCCTGACCCACTAAATTGATTATGTAAC 19683
QY 204 ----- 204
Db 19684 CTGCTAATGTGTATGAATCTTAAATTTGAAAATTAGTGACATAGTACATATTGTTTCATC 19743
QY 204 ----- 204
Db 19744 TTTGAGTGCTTTTAAATGTATACTTTAAGGTATAGAGGTTTCATTATACAGTGTAT 19803
QY 204 ----- 204
Db 19804 TTGTGTGCTGTTTAAACATATACAATATCCTAGCTTATTTCTAAAGTCAAACTTTAA 19863
QY 205 ----- 208
Db 19864 AATTTCATGGCTTATATGAATTTCATAGTTTCCCTTGACCTTCTTTTCAGAGGGAATGA 19923
QY 208 pLYsAspPheThrLeuAsnAspPheGlyPheMetIlePheHisSerProTyrCysIysLe 228
Db 19924 TAAAGATTTTACCTTGAATGATTTGGCTTCATGATCTTTCACCTCACCATATTTGTAAC 19983
QY 228 uValGlnIysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsnAspGlnAsnArgAs 248
Db 19984 GGTTCAGAAATCTCTAGCTCGGATGTGCTGTAATGACTTCTTAATGACCAAGATAGAGA 20043
QY 248 pLYsAsnSerIleTyrSerGlyLeuGlnAlaPheGly----- 260
Db 20044 TAAAAATAGTATCTATAGTGGCTGGAAGCCTTTGG-GTAAGAGGAGCTATATGAGTTT 20102

QY 260 ----- 260
Db 20103 TTTCCCTCTATATAGACATTTTAAATATCTGTTAAAGCTGTATTTGTACAGACCTGAG 20162
QY 260 ----- 260
Db 20163 AAATTGAGAGTCAGAGAATCTTAGAAGTCATCCAGTCTAATCTGTGTCTCAGTCAGT 20222
QY 260 ----- 260
Db 20223 GAAGAATCTAAGTCCAGAGAGGTGTAGTTAACATGCACAAATTTCTTAGACATTTCTAT 20282
QY 260 ----- 260
Db 20283 TCAGATTTTCTGATTTATTTCTTTCAGCTCCATTCATGTGTACAGATAAAGTAAGTCA 20342
QY 260 ----- 260
Db 20343 CAAGGGCTATATTCATACAGCAGCCTCTTAACCTTACCTCTCTCAGCACCCCTGCC 20402
QY 260 ----- 260
Db 20403 CCCATGCCCTTTTCCATCCTGCACACTGCCACAGCTAAAGTCAGCTTTGTACTCCACT 20462
QY 260 ----- 260
Db 20463 GTCTTTTCTCACTTTAGGCTCCCTAGCATGTATGTGTCAACTCGTTCTGTCTCTC 20522
QY 260 ----- 260
Db 20523 CCTGTCTCTTGTGTCTCTTCTCTATCTGATAAATTATACTTGACTTTAAACTT 20582
QY 260 ----- 260
Db 20583 GGCTCCTGTAATACCATGACTTTTCTAACTAATAAACAATTATGAGACTTGAATAGT 20642
QY 260 ----- 260
Db 20643 ATTCTATTCAAGTGATGAATATTCAGTTGATGAAATATCTATTCAATTGAGCCAATATA 20702
QY 260 ----- 260
Db 20703 AGTGAATATAAATAAAGCTACAGTGCCTTTTAACTATTCAAAATCAAGCAGCCTTA 20762
QY 260 ----- 260
Db 20763 ACTTGATTATGAAAACCTTTTGAGAAAAAGAACCATATATACAACTGTTATGATTTCTA 20822
QY 260 ----- 260
Db 20823 TAGCAATTAGATTGCTGCTACTTGCTTTTAATATAAGAAAAACAATTATATACACTTA 20882
QY 260 ----- 260
Db 20883 AAGATTTGAATCCTAATTAGGCTGCTGTTTAGTGTAAATAAAAAACATAGGCTTTAAACAC 20942
QY 261 ----- 271
Db 20943 TGTAAACTGTAAATAAATCTTTCAGGGATGTTAAATTAGAAGACACCTACTTGTAGTAG 21002
QY 271 gAspValGlnIysAlaPheMetLysAlaSerSerGlnLeuPheSerGlnIysThrIysAl 291
Db 21003 AGATGTGAGAGGCAATTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAGGC 21062
QY 291 aSerLeuLeuValSerAngIlnAsnGlnIysAsnMetIyrThrSerSerValTyrGlySerle 311
Db 21063 ATCTTTACTTGTATCAAAATCAAAATGAAATATGTACACATCTTCAGTATATGTTCCT 21122
QY 311 uAlaSerValLeuAlaGln----- 317
Db 21123 TGCATCTGTTCTAGACAGATAAGTATAAATTCACCTACTACTTAACCTCCCTTATTTGG 21182
QY 317 ----- 317

Db 21183 GAGATGTTAGATTCTTAAGACCAAACTAGTGTCAAGCATGTGGTAGATCACAGAA 21242
QY 317 ----- 317
Db 21243 AATTTTATCTGAGGCTCTCTAATCTGCTAATGTCCATTGACTTGAAAAGATGATGGGT 21302
QY 317 ----- 317
Db 21303 GAGGCTACAGTCTTCCAGAAGTATTGTTAATTTCATCTGGCTTCCCTGGCTTCTGTT 21362
QY 317 ----- 317
Db 21363 TTCATGTTTTTAATTCTTGACCTACAGTTGAACCAATAATACCTGGTTGATGAAGTAA 21422
QY 317 ----- 317
Db 21423 CTTGTTTGTGGCATGACTTTCACAAGCTCTGTTCATCCCCACAGAATGAAAACTCACAT 21482
QY 317 ----- 317
Db 21483 GCTGCAATATTAATAAATAAGTTAATTCCTTACATATTAACAATTGAGTTAGATC 21542
QY 317 ----- 317
Db 21543 CTTAAAACTTTAAGTTAGATTCTACTTTTACTTATAGCCTAAATTTTATGTACTTTT 21602
QY 317 ----- 317
Db 21603 ATAGCTTCCACACGCTGTAGCTTTGGATCAGTTAAACTTCTGAACCTATTGTTACACCCCT 21662
QY 318 -----TyrSerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySe 335
Db 21663 ACATAGGTACTCACTCAGCAATTAGCAGGGAGAGAATTGAGTGTCTTTCTTATGGTTC 21722
QY 335 rglyLeuAlaAlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrPro----- 352
Db 21723 TGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTAAAGTG 21782
QY 352 ----- 352
Db 21783 CTGAATCTTTCACAAGAATGTATTGAGAACTGAGTCCAGGCAAGTGCGTCAACCCGT 21842
QY 352 ----- 352
Db 21843 AATCCACAGAGTTTGGAGGCGGAGCGGACAGATCACCTGAGTCAAGAGTTCAGAGACC 21902
QY 352 ----- 352
Db 21903 AGTCTGCTAACATGCGCTGAACCCCATCTCTACTAATAAAATACAAAAATTAGCCAGGTGA 21962
QY 352 ----- 352
Db 21963 GGTGTGCATGCTGTAGTCTTAGCTACTTGGGAGGCTGAAGTAGGAGAATCACTTGAAT 22022
QY 352 ----- 352
Db 22023 CCAGGAGAGGAGGTTGTGTGAGCCAAAGATCACACCACCTGTGCTCCAGCCTGGGTGACA 22082
QY 352 ----- 352
Db 22083 GAGCGAGACTCTGTCAAAAAAAAAAAAAAAAAATGTAATTGAGAACTACTCTGGGGAAAGTT 22142
QY 352 ----- 352
Db 22143 GATTTAGCAGTCTTCTCAAGTAGCACTGAATCTGTCCACAGATCATTAATATTTT 22202
QY 352 ----- 352
Db 22203 AGTCTTCACTACTCTTTCAGTAGGTTTTTACTCTGTGCCCTAATAATCTATCCAAAAA 22262
QY 352 ----- 352

Db 22263 AAAAAAAAAATTCTACCTTATCTGGATTAAGAGTAGGACTAAGTTATCTAATTTTTATAGG 22322
QY 352 ----- 352
Db 22323 CTTATGCTCTGGCTATATTTAAAGTCACTTTTGTGCTTCCCTGAGCAGAAAGACCA 22382
QY 352 ----- 352
Db 22383 AAATGTAGAGTAATAACTGATGA AAACTTGACATTACTTTTAAATATATACCATGGGCCA 22442
QY 352 ----- 352
Db 22443 GGTGCAATGGCTCACACCTATTAATCCCAACACTTCAGAGGCTGAGGTGGAGAGATTGCT 22502
QY 352 ----- 352
Db 22503 TGAGGCCAGATGTTCAAGGCCAACCTGAGCAACATAGTAGAGACCCCATCTCTATPAAAAA 22562
QY 352 ----- 352
Db 22563 TAATAAAAAATAAATAATTAATTAACATGAGTTAATTGTAGACAAAGTTATTATAGTTCAA 22622
QY 352 ----- 352
Db 22623 ATTATGCCGTGTTTCTTAACCTTGCTCTAGTGGCAGATACTCAATTAATAGATTCTAGCTGA 22682
QY 352 ----- 352
Db 22683 CATCATAGAGATTGTCAATAAGTATCATCTTATCTTTAATAATCAGTAGCCAGTA 22742
QY 352 ----- 352
Db 22743 GTTTTAATGA AAAATGA AAAAGTTGTTTGCCTCATTTGGCAACATTTTACTTAGGCTTCTT 22802
QY 353 -----glySerA 355
Db 22803 TTGCACATGATTTTTCAAAAAAATCTTTTAATGTTGAATTATTCATATTTTAGGGTCTG 22862
QY 355 lal euAspLysileThrAlaSerLeuCyAspLeuLysSerArgLeuAspSerArgThrG 375
Db 22863 CTCTTGATTAATAATAACAGCAAGTTTATGTGATCTTAATCAAGGCTTGATTCAGAAACTG 22922
QY 375 l yValAlaProaspValPheAlaGluAsnMetLysLeuArgGluAspThrHis---- 393
Db 22923 GTGTGCACGAGATGTCTTCGCTGA AAAACATGAAGCTCAGAGAGGACACCCATCATTTTGG 22982
QY 393 ----- 393
Db 22983 GTAAAAATATTAATGTCTTTAAGTTAACCCATTTGGAGGGCTGATATCATTTAAGGATG 23042
QY 393 ----- 393
Db 23043 CTACATATACGATAAGATATCAAGACTTTACTCAGTACTAATCTGATGTCAATGAAAAAT 23102
QY 393 ----- 393
Db 23103 TATTGGATATGA AACTTATCTTAGCTTTATTAACAGATGAATGTATATCA TAACT 23162
QY 394 -----LeuValAsnTyrIleProGlnGlySerIleAspS 405
Db 23163 AATTGTAGATATCTCTCCCTTCTTACTTAGTCAACTATATCCCCAGGGTTCAATAGATT 23222
QY 405 erLeuPheGluGlyThrTyrPyrLeuValArgValAspGluLysHisArgArgThrTyrA 425
Db 23223 CACTCTTTGAAGGAACGTGTACTTAGTTAGGGTGATGA AAAAGCACAGAAAGAACTTACG 23282
QY 425 laArgArgProThrProAsnAspAspThrLeuAspGluGlyValGlyLeuValHisSerA 445
Db 23283 CTGGCGTCCCACTCCAAATGATGACACTTTGATGAAGGAGTAGACTGTGTGATTCAA 23342
QY 445 snl leAlaThrGluHisIleProSerProAlaLysLysValPro 459
Db 23343 ACATAGCAACTGAG-----GTAAATPAAAAAGATTCCC 23374

RESULT 5
US-10-193-295-3
; Sequence 3, Application US/10193295
; Patent No. 6620608
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: C1001195DIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; PRIOR FILING DATE: 2002-07-12
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 28001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(28001)
; OTHER INFORMATION: n = A,T,C or G
US-10-193-295-3

Alignment Scores:
Pred. No.: 5.76e-92 Length: 28001
Score: 839.50 Matches: 302
Percent Similarity: 19.23% Conservative: 1
Best Local Similarity: 19.16% Mismatches: 3
Query Match: 33.43% Indels: 1271
DB: 4 Gaps: 6

US-10-622-516-2 (1-478) x US-10-193-295-3 (1-28001)

QY 150 GlyLeuArgGlyThrHisMetGlnHisAlaTyrAspPheTyrLysProAspMetLeuSer 169
Db 18664 GGGCTTCGTGGGACACATATGCAACATGCCCTATGATTTTACAAAGCCTGATATGCTATCT 18723
QY 170 GluTyrProIleValAspGlyLysLeuSerIleGlnCysTyrLeuSerAlaLeuAspArg 189
Db 18724 GAATATCTTATAGTAGATGGAATACTCTCCATACAGTGTCTACCTCAGTGCATTAGACCGC 18783
QY 190 CysTyrSerValTyrCysLysLysIleHisAlaGlnTyrGln-Lys----- 204
Db 18784 TGGTATTTCTGTCTACTGCAAAAAGATCCATGCCAGTGGCAGAAAGGTAAGTTTACCGA 18843
QY 204 ----- 204
Db 18844 TTTTCCTTGGTTTGTATGAGTTGAGAGCAGTCTAATGTAAGTATCTTGGTAGGC 18903
QY 204 ----- 204
Db 18904 AACTACTTTGTGGCATTCTTCATTTAATATCCTTTACCATTAATTCCTCATTCACCA 18963
QY 204 ----- 204
Db 18964 ACAACATTTTCCCATAGTTTCTGGGAAAGTGTAATTACTAGAGAAGGTAACCTTTGGA 19023
QY 204 ----- 204
Db 19024 CTGAGGTGTATCTCTGCAAAATAATTAGGTCGGTTTACCCCTGTGAAGAAATCAAGT 19083
QY 204 ----- 204
Db 19084 GGAAGAAAGAAAGTAAGTTGTAATTTGTTCATCTTTGAGAGAGGTATTTTAACAAGGT 19143
QY 204 ----- 204
Db 19144 TTGGACTACAGCTGTGATTCAGGGAAGCTAATGAATAAATACTAAGTATCTTAC 19203

QY 204 ----- 204
Db 19204 CCCAAAAATAATCTTTTGGCACTTGACCTGTGAATTTGTATTTTCTACTGTATC 19263
QY 204 ----- 204
Db 19264 ATTAATCTGAAATTTGTGAGGCACTGAAGAGACAGTATTGAGTTAATGCTATCAATA 19323
QY 204 ----- 204
Db 19324 CACATTATTACATAAAGTATACCTTTTCTGTAGTCCAATTGCTTTTAGAGGTATGA 19383
QY 204 ----- 204
Db 19384 GAAGGGGTTAAAAATCATATTCAATGACAATAATCAGTGAATTTAGTCGCTGGAATAAG 19443
QY 204 ----- 204
Db 19444 AAGCATTCTTGCAGTATATATTAAACAGATAGTGTCTTCTAATCTTTTATTAGGACCC 19503
QY 204 ----- 204
Db 19504 ACAGTAAGAAGTACAGTGTATACATTGTATGTATGCGACGACTGAACAAAAATGTCAATGA 19563
QY 204 ----- 204
Db 19564 CATTACTTACCCTTGCTGCAAGTTATTCAAGTTTGCTATTCTTCTACTGCATTTGTCTTTT 19623
QY 204 ----- 204
Db 19624 TAAATACTCTTTTATTTAAAAAATACTAATCCGACCCCAATAATTGATTATGTAAC 19683
QY 204 ----- 204
Db 19684 CTGCTAATGTGTATGAATCTTAAATTGAAAAATTAGTACATAGTACATATTGTTTCATC 19743
QY 204 ----- 204
Db 19744 TTTGAGTGTCTTTTAAATGTATACCTTAAAGGTATAGAGAGGTTTCATTATACAGTGTAT 19803
QY 204 ----- 204
Db 19804 TTGTGGTTGCTGTTTAAACATATACAATAATCCTAGCTTATTTCTAAAGTCAAACTTTAA 19863
QY 205 ----- 208
Db 19864 AATTTCATGGCTTATATGAATTTCATAGTTTCCCTTGGACTTCTCTTTCAGAGGGAATGA 19923
QY 208 PLYSAspPheThrLeuAsnAspPheGlyPheMetIlePheHisSerProTyrCysLysIle 228
Db 19924 TAAAGATTTTACCTTGATGATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACT 19983
QY 228 uValGlnLysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsnAspGlnAsnArgAs 248
Db 19984 GGTTCAGAAATCTCTAGCTCGGATGTTGCTGAATGACTTCTTAATGACCAGAAATGAGA 20043
QY 248 PLYSAsnSerIleTyrSerGlyLeuGlnAlaPheGly----- 260
Db 20044 TAAAAATAGTATCTATAGTGGCCCTGGAAGCCTTTGG- GTAAGAGAGCTATTATGAGTTT 20102
QY 260 ----- 260
Db 20103 TTTCTTCTATATTAGAGCAATTTTAATATCTGTTAAGCTGTATTGTACAGACCTGAG 20162
QY 260 ----- 260
Db 20163 AAATTGAGAGTCAGAGAATCTTAGAAGTCATCCAGTCTAATCTGTGTCTCAGTCAGT 20222
QY 260 ----- 260
Db 20223 GAAGAACTTAAGTCCAGAGAGGTGTAGTTAACATGCACAAATCTTTAGACATTTCTAT 20282
QY 260 ----- 260

```
Db 20283 TCAGATTTTCTGATTATTCTTTTCAGCTCCATTCATGTGTGTCACGATAAAGTAACTGCA 20342
QY 260 ----- 260
Db 20343 CAAGGCCATAATTCTACTACAGCAGCCTCTTAACCTCCTTACCCTCTCTCAGACCCCTGCC 20402
QY 260 ----- 260
Db 20403 CCCATGCCCTTTCCATCCCTGCACACTGCCACAGCTAAAGTCAGCTTTTGTACTCCACCT 20462
QY 260 ----- 260
Db 20463 GTCTTTTCTCACTTAGGCTCCCTAGCANGCTATGTGTGTTCAACTGTTCTGTTTCTC 20522
QY 260 ----- 260
Db 20523 CCTGTGTCCTGTGTGTGTCCTTTTCTCTATCTGATAAAATTATATCTTGACTTTTAAACTT 20582
QY 260 ----- 260
Db 20583 GGCTCCTGTAATACCATGACTTTTCTAATAATAACATTTATGTGACTTGAATAAGT 20642
QY 260 ----- 260
Db 20643 ATTCTATTCACTGTGATGAATATTCAGTTGATGAATATTTCTATTCATTGAAGCAATATA 20702
QY 260 ----- 260
Db 20703 AGTGAATATAAATATAAAGCTACAGTGCCTTTTAACTATCAAATCAAGCAGCTTA 20762
QY 260 ----- 260
Db 20763 ACTTGATTATGAAAACTTTTGAGAAAAAGAACATATATATACAACGTGTTATGATTTCTA 20822
QY 260 ----- 260
Db 20823 TAGCAATTAGATTGCTGCTACTTGCTTTTAATAATGAGAAAACAATTATATACACTTA 20882
QY 260 ----- 260
Db 20883 AAGATTGAAATCCTAATTAGGCTGCTGTTTACGTAAATAAAAACATAGGCTTTAAACAC 20942
QY 261 -----AspValLysLeuGluAspThrTyrPheAspAr 271
Db 20943 TGTAAACTGTAAATAAATATCTTTCAGGGATGTTAAATTAGAGACACCTACTTTGATAG 21002
QY 271 gAspValGluYsaAlaPheMetLysAlaSerSerGluLeuPheSerGlnLysThrLysAl 291
Db 21003 AGATGTGAGAGGCATTTATGAGGCTAGCTCTGAACCTTTCAGTCAGAAAAACAAGGC 21062
QY 291 aSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyrThrSerSerValTyrGlySerLe 311
Db 21063 ATCTTTACTTGTATCAAAATCAAAATGGAATATGTACACATCTTCAGTATATGTTCCCT 21122
QY 311 uAlaSerValLeuAlaGln----- 317
Db 21123 TGCATCTGTTCTTACACAGTAACTAAATTTCACCTACTACTTAACCTCCCTTAATTGG 21182
QY 317 ----- 317
Db 21183 GAGATGTTAGATTCTTAAGACCAATCTAGTGTCAAGCATGTTGGTGATGATCACAGAA 21242
QY 317 ----- 317
Db 21243 AATTATCTTGAGGCTCTCTAATCTGCTATGTCCATTGACTTGAAAAGATGATGGGTT 21302
QY 317 ----- 317
Db 21303 GAGGCTACAGTTCTTCCAGAACTATTGTTAATTTCATACCTGGCTTCCCTGCTTCTGTT 21362
QY 317 ----- 317
```

```
Db 21363 TTCATGTTTTTTTAATTCCTTGACCTACAGTTGAACCATAAATACTGTTGATGAAGTAA 21422
QY 317 ----- 317
Db 21423 CTTGTTTGTGGCATGACTTTTCAAGACTGTGTCAATTCCTCCACAGATGAAAACTCACAT 21482
QY 317 ----- 317
Db 21483 GCTGCAATATTAATAACTAAGTTATATCCCTACTGCAATATTAACACTTGAAGTTAGATC 21542
QY 317 ----- 317
Db 21543 CTTAAACTTTAAGTTAGATTCCTACTTTACTTATAGCCTAAATTTTATTTGCTACTTTT 21602
QY 317 ----- 317
Db 21603 ATAGCTTCCCAACGCTGTAAGCTTTGGATCAGTTAAACTTGTGAACCTATTTGTACACCT 21662
QY 318 -----TyrSerProGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySe 335
Db 21663 ACATAGTACTCACTCAGCAATTAGCAGGGAAGAAATGGAGTGTCTTCTTAAGTTC 21722
QY 335 rgLYLeuAlaAlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrPro----- 352
Db 21723 TGGTTGGCTGCCACTGTGACTCTCTTAAAGTCACACAAGATGCTACACCGGTTAAGTG 21782
QY 352 ----- 352
Db 21783 CTGAATCTTTCAACAAGAATGTAATTGAGAACTGAGTCCAGGCACAGTGGCTCACACCCGT 21842
QY 352 ----- 352
Db 21843 AATCCAGCAGTTTGGAGGCCGAGCGGCGAGATCACCTGAGGTCAGAGTTGAGAGACC 21902
QY 352 ----- 352
Db 21903 AGTCTGCTAACATGGCTGAAAACCCCATCTCTACTAAATAACAAAAATTAGCCAGGTGA 21962
QY 352 ----- 352
Db 21963 GGTGTGCATGCTGTAGTCTTACTACTTTGGGAGGCTGAAGTAGAGAAATCACTTGAAT 22022
QY 352 ----- 352
Db 22023 CCAAGAGAGGAGGTTGTGTGAGCCAAAGATCACACCACTGTGCTCCAGCTGGGTGACA 22082
QY 352 ----- 352
Db 22083 GAGCGAGACTCTGTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 22142
QY 352 ----- 352
Db 22143 GATTTAGCAGTCTTCTCAAGTGAACACCTGAATCTGTCCACAGATCATTAATAATTTT 22202
QY 352 ----- 352
Db 22203 AGTCTTCATTACTTCTTTCAGTAGGTTTTTACTCTCTGCCCCATAAATACTATCCAAAAA 22262
QY 352 ----- 352
Db 22263 AAAAAAAAAATTCTAAGTTATCTGATATAAAGATAAGACTAAGTTATCTAATTTTATAGG 22322
QY 352 ----- 352
Db 22323 CTTATGCTCTGGCTATATTTAAGTCACTTTTGTGCTTTCCTGAGCAGAAAGACAA 22382
QY 352 ----- 352
Db 22383 AAATGTAGAGATAAACTGATGAAAACTTGACATTAATTTTAAATTATACCATGGGCCA 22442
QY 352 ----- 352
Db 22443 GGTGCAATGGCTCACACCTATAATCCCAACACTTCAGGAGGCTGAGGTGGAGATTGCT 22502
```



```

QY      352 ----- 352
Db      22503 TGAGGCCAGATGTTCAAGGCCAACCTGACACATAGTGACACCCCATCTCTATAAAAA
QY      352 ----- 352
Db      22563 TAATAAAATAAATAATATATACCATGATTAAATTGAGCAAGTTATTATAGTTTCAA
QY      352 ----- 352
Db      22623 ATTATGCCCTGTTCCCTAACTGTCTAGTGCGACAGATACTCAATAATAGATTCTAGTCTGA
QY      352 ----- 352
Db      22683 CATCATAGAGATTGTCAAAATAGGTATCATCTTATCTTTAACTAATCAGTAGCCAGTA
QY      352 ----- 352
Db      22743 GTTTAAATGAAAAGTTGTTTGCCCTCATTTGGCAACATTTACTTAGGCTTCTT
QY      353 -----GlySerA 355
Db      22803 TTGGACATGATTTTCAAAAAAATCTTTAATGTTGAATTAATCACTATTTTAGGGTCTG
QY      355 1aLeuAspLysIleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrG 375
Db      22863 CTCTGATAAAATAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 22922
QY      375 1yValAlaProAspValPheAlaGluAsnMetLysLeuArgGluAspThrHis---- 393
Db      22923 GTGTGGCACCAAGATGTCTCGCGAAAAACATGAAGCTCAGAGAGACACCCCATCATTTGG 22982
QY      393 ----- 393
Db      22983 GTAAAAATATTAATGTCTTAAAGTTAACCCATTGGAGGCTGATATCATTAAGATG 23042
QY      393 ----- 393
Db      23043 CTACATATACGATAAGATATCAAGACTTTACTCAGTACTAATCTGATGTCAGTGAAT 23102
QY      393 ----- 393
Db      23163 AATTGTAGATATTCTCTCCCTTCTCTTAACTGCACTATATCCCCAGGGTCAATAGATT 23222
QY      394 -----LeuValAsnTyrIleProGlnGlySerIleAspS 405
Db      23163 AATTGTAGATATTCTCTCCCTTCTCTTAACTGCACTATATCCCCAGGGTCAATAGATT 23222
QY      405 erLeuPheGluGlyThrTyrPheLeuValArgValAspGluLysHisArgArgThrTyrA 425
Db      23223 CACTCTTGAAGAACGTGTACTTAAGGTGATGAAAGCACAGAACTTACG 23282
QY      425 1aArgArgProThrProAsnAspThrLeuAspGluGlyValGlyLeuValHisSerA 445
Db      23283 CTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAGAGTAGAAGTGTGCATTCAA 23342
QY      445 snIleAlaThrGluHisIleProSerProAlaLysValPro 459
Db      23343 ACATAGCAACTGAG-----GTAATATAAAGAGTTCCC 23374

```

```

RESULT 6
US-09-370-838-245
; Sequence 245, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1

```

```

; CURRENT APPLICATION NUMBER: US/09/370, 838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 245
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105)
; OTHER INFORMATION: n=A,T,C or G
US-09-370-838-245

```

```

Alignment Scores:
Pred. No.: 2,45e-86 Length: 615
Score: 771.00 Matches: 148
Percent Similarity: 99.33% Conservative: 0
Best Local Similarity: 99.33% Mismatches: 0
Query Match: 30.70% Indels: 1
DB: 4 Gaps: 0

```

US-10-622-516-2 (1-478) x US-09-370-838-245 (1-615)

```

QY      3 GlySerLeuProLeuAsnAlaGluAlaCysTyrProLysAspValGlyIleValAlaLeu 22
Db      108 GGATCACTTCCTTGATGACAGAAAGCTTGCTGGCCAAAGATGTGGAAATGTGCCCTT 167
QY      23 GluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAspGlyVal 42
Db      168 GAGATCTATTTCCTTCTCAATATGTGTGATCAAGCAGAGTTGGAAAAATATGATGGTGT 227
QY      43 AspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThrAspArg 62
Db      228 GATGCTGAAAGTATACCAATGGCTTGGCCAGGCCAAGATGGGCTTCTGCACAGATAGA 287
QY      63 GluAspIleAsnSerLeuCysMetThrValAlaGlnAsnLeuMetGluArgAsnAsnLeu 82
Db      288 GAGATATTAATCTCTTGGCATGACTGTGTTCAAGATCTTATGAGAGAAATAACCTT 347
QY      83 SerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLysSerLys 102
Db      348 TCCTATGATTGCATTTGGCGGCTGGAAGTTGGAACAGACATCATCGACAATCAAG 407
QY      103 SerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIleGluGly 122
Db      408 TCTGTGAAGACTAATTGTATGACAGCTGTTGAAGAGTCTGGGAATACAGATATAGAAAG 467
QY      123 IleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAla-ValAsnTr 142
Db      468 ATGACACAACATAATGATGATGAGGACAGCAGCTGTCTTCAATGCTTGTAACTG 527
QY      142 PileGluSerSerSerTrpAspGly 150
Db      528 GATTGAGTCCAGCTCTTGGGATGGA 552

```

```

RESULT 7
US-09-370-838-128
; Sequence 128, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370, 838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323

```

; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for windows Version 3.0
; SEQ ID NO 128
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-370-838-128

Alignment Scores:
Pred. No.: 5,09e-78 Length: 500
Score: 703.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 28.00% Indels: 0
DB: 4 Gaps: 0

US-10-622-516-2 (1-478) x US-09-370-838-128 (1-500)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 102 ATGCTCGATCACTTCCTTGAATGCAGAAAGCTTGCTGCCCAAAAGATGGGAATTGTT 161
QY 21 AlaLeuGluIleTyRpheProSerGlnTyRValAspGlnAlaGluLeuGluLysTyRAsp 40
Db 162 GCCCTTGAGATCTAATTTCTCTCAATATGTTGATCAAGCAGAGAGTTGAAAAATATGAT 221
QY 41 GlyValAspAlaGlyLysTyRThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 222 GGTGTAGATGCTGAAAGTATACCATGGCTTGCGCCAGGCCAAGATGGGCTTCTGCACA 281
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValAlaGlnAsnLeuMetGluArgAsn 80
Db 282 GATAGGAAGATATTAACTCTCTTGCAATGACTGTGCTTCAGAACTTATGAGAGAAAT 341
QY 81 AsnLeuSerTyRAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 342 AACCTTCCCTATGATGTCATTTGGCGCGCTGGAAGTGGAAACAGAGACATCATGCACAA 401
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 402 TCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGAATACAGATATA 461
QY 121 GluGlyIleAspThrThrAsnAlaCysTyRGlYglYThr 133
Db 462 GAAGGATCGACACAACTAATGCAATGAGAGCACACA 500

RESULT 8

US-09-401-064-187
; Sequence 187, Application US/09401064
; Patent No. 6623923
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeline Joy
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C2
; CURRENT APPLICATION NUMBER: US/09/401,064
; CURRENT FILING DATE: 1999-09-22
; NUMBER OF SEQ ID NOS: 371
; SOFTWARE: FastSeq for windows Version 3.0
; SEQ ID NO 187
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-401-064-187

, Alignment Scores:

Pred. No.: 6.84e-76 Length: 506
Score: 686.00 Matches: 130
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.32% Indels: 0
DB: 4 Gaps: 0

US-10-622-516-2 (1-478) x US-09-401-064-187 (1-506)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db 117 ATGCTCGATCACTTCCTTGAATGCAGAAAGCTTGCTGCCCAAAAGATGGGAATTGTT 176
QY 21 AlaLeuGluIleTyRpheProSerGlnTyRValAspGlnAlaGluLeuGluLysTyRAsp 40
Db 177 GCCCTTGAGATCTAATTTCTCTCAATATGTTGATCAAGCAGAGAGTTGAAAAATATGAT 236
QY 41 GlyValAspAlaGlyLysTyRThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 237 GGTGTAGATGCTGAAAGTATACCATTTGGCTTGCGCCAGGCCAAGATGGGCTTCTGCACA 296
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValAlaGlnAsnLeuMetGluArgAsn 80
Db 297 GATAGGAAGATATTAACTCTCTTGCAATGACTGTGCTTCAGAACTTATGAGAGAAAT 356
QY 81 AsnLeuSerTyRAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 357 AACCTTCCCTATGATGTCATTTGGCGCGCTGGAAGTGGAAACAGAGACAAATCATGCACAA 416
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 417 TCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTGAAGAGTCTGGGAATACAGATATA 476
QY 121 GluGlyIleAspThrThrAsnAlaCysTyR 130
Db 477 GAAGGATCGACACAACTAATGCAATGAGAGCACACA 506

RESULT 9

US-09-306-595C-1
; Sequence 1, Application US/09306595C
; Patent No. 6284506
; GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: OJIMA, Kazuyuki
; APPLICANT: SETOGUCHI, Yutaka
; TITLE OF INVENTION: ISOPRENOL PRODUCTION
; FILE REFERENCE: ISOPRENOL PRODUCTION
; CURRENT APPLICATION NUMBER: US/09/306,595C
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 98108210
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4775
; TYPE: DNA
; ORGANISM: Phaffia rhodozyma
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1239)..(1240)
; OTHER INFORMATION: EXPERIMENTAL
; NAME/KEY: exon
; LOCATION: (1305)..(1361)
; NAME/KEY: intron
; LOCATION: (1362)..(1504)
; NAME/KEY: exon
; LOCATION: (1505)..(1522)
; NAME/KEY: intron
; LOCATION: (1523)..(1699)
; NAME/KEY: exon
; LOCATION: (1700)..(1826)
; NAME/KEY: intron
; LOCATION: (1827)..(1920)

```

; NAME/KEY: exon
; LOCATION: (1921) .. (2277)
; NAME/KEY: intron
; LOCATION: (2278) .. (2351)
; NAME/KEY: exon
; LOCATION: (2352) .. (2409)
; NAME/KEY: intron
; LOCATION: (2410) .. (2497)
; NAME/KEY: exon
; LOCATION: (2498) .. (2504)
; NAME/KEY: intron
; LOCATION: (2505) .. (2586)
; NAME/KEY: exon
; LOCATION: (2587) .. (2768)
; NAME/KEY: intron
; LOCATION: (2769) .. (2851)
; NAME/KEY: exon
; LOCATION: (2852) .. (2891)
; NAME/KEY: intron
; LOCATION: (2892) .. (2985)
; NAME/KEY: exon
; LOCATION: (2986) .. (3240)
; NAME/KEY: intron
; LOCATION: (3241) .. (3325)
; NAME/KEY: exon
; LOCATION: (3326) .. (3493)
; NAME/KEY: intron
; LOCATION: (3494) .. (3601)
; NAME/KEY: exon
; LOCATION: (3602) .. (3768)
; NAME/KEY: polyA site
; LOCATION: (4043) .. (4044)
US-09-306-595C-1

```

Alignment Scores:

Pred. No.:	1,36e-61	Length:	4775
Score:	585.00	Matches:	181
Percent Similarity:	37.15%	Conservative:	82
Best Local Similarity:	25.56%	Mismatches:	129
Query Match:	23.30%	Indels:	318
DB:	3	Gaps:	15

US-10-622-516-2 (1-478) X US-09-306-595C-1 (1-4775)

[illegible]

Db	2072	TATCGACTCCAGATGCGCTGCTACGGTCTTACCGCGCCCTGTTCATATGCCGTCACTG	2131
QY	142	pIleGluSerSerSerTrpAspGlyLeuArg-----	152
Db	2132	GATCGAGTCATCCTCTTGGGACGGAG- AAATGCCATTGTCTTTCGGGAGACATTGCCA	2190
QY	152	-----	152
Db	2191	TCTACGCCGAGGTCGTGCCGACCTGCGGAGGTGCTGTGCTTGCGCCATCTCATCG	2250
QY	152	-----	152
Db	2251	GACCCGACGCTCCCGTCGTCTTCGAGCGTAGTTCCAATCCGTCATTTTCTTCCACGGCA	2310
QY	153	-----GlyThrHisMe	156
Db	2311	GCGGCTGAACAACCCCTTATCCGTCATTCTCATCAATCTAGCCGTCACCGAAACTTCAT	2370
QY	156	tGlnHisAlaTyrAspPheTyrLysProAspMetLeuSer-----	169
Db	2371	GACCAACGCTTGGACTTCTACAAGCCTAATCTTCTTCGTATGTTCAAATTGTGAAGTT	2430
QY	169	-----	169
Db	2431	TGCGCTTGAGAGTCTTACACTAATTCGGGGTGCATCCTTGAATCGTTGTTGC	2490
QY	169	-----	169
Db	2491	TTTATAGTAATACGTTGCTCGCGCACCTCCTATATTAGTTTGTGATCAAAATATTGTC	2550
QY	170	-----GluTyrProIleValAspGlyLysLeuSe	179
Db	2551	CATTGAATTAACTGTGAACCTTCTCCTCCAATAG- CCCATTTGCGATGAGACCTTCTC	2609
QY	179	rIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIleHi	199
Db	2610	CGTCACTTCCCTACGTCACGCCCATTTGACAAAGGCTTGAAGCTTACCGAACAAAGTATGC	2669
QY	199	sAlaGlnTrpGlnLysGluGlyAsnAspLys-----	209
Db	2670	CAAGCGATT-----GGAGGACCCAAAGACTTAACGGTGTCAACCAACGACACACCGA	2720
QY	210	-----AspPheThrLeuAsnAspPheGlyPheMetIlePheHis-----	222
Db	2721	GGTTGCCGGTGTCAGTGCTGCGTTCGATTACCTTTGTTCACAGGTAAGCGTCATC	2780
QY	223	-----SerProTyr-----	225
Db	2781	TTCTGTATTCTCCTTAATTCACACCGATCAACGAGTTAATTCGTGCATCATATTATCT	2840
QY	226	-----CysLysLeuValGlnLysSerLeuAlaArgMet-----	236
Db	2841	TGTTGGAACAGTCTTACGGAAGACAGTTGTCAAAAGGCCACGGCGACTTGTAAACAGT	2900
QY	236	-----	236
Db	2901	CTTTTGTAACTTTAGCTTGCAGATAAAAATTGAGTTTCTGTACTCATTTATTAT	2960
QY	237	-----LeuLeuAsnAspPheLeuAsnAspGlnAsnArgAs	248
Db	2961	GCATCTCTTGAATCACTTATCTAGTTGTACATGACTTCGAAACAACCCCAAC-----	3015
QY	248	plyAsnSerIleTyrSerGlyLeuGluAla--PheGlyAspValLysLeuGluAspTh	267
Db	3016	---GACCCGGTTTGTCTGAGGTGCCAGCCGAGCTTGCTACTTTGGACATGAGAAAAG	3071
QY	267	rTyrPheAspArgAspValGluLysAlaPheMetLysAlaSerSerGluLeuPheSerG1	287
Db	3072	TCTTTCAGACAAGATGTCGAGAAATCTGTGATGTGCTCCCAAGTCTTCTTCAACAA	3131
QY	287	nLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyrThrSerSerVa	307


```
Db 3132 GCAGGTGAGCCTGGAATGACCAACCGTCCGACAGCTCGGAACCTGTACACCGCCTCTCT 3191
Qy 307 1TyrGlySerLeuAlaSerValIleuAlaGlnTyrSerProGlnGlnLeu----- 323
Db 3192 CTTGGTGTCTCTCGAAGTTGTTCTTAATGTTCCTGTGACGAGCTCGAAGTCTTGA 3251
Qy 323 ----- 323
Db 3252 TCTCTATCCCAATCATCTCTTCTTATCAATTGAACCTGAACTCTTTCTTTAATGCTGGC 3311
Qy 324 -----AlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaA 339
Db 3312 TTTCTCTTGAAACAGGTGCGCAAGCGCATGTCTCTACGCGCTACGGATCTGGAGCTGTG 3371
Qy 339 1aThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysI 359
Db 3372 CTTCTTTCTAATGCTCTTAAGGTCAAGAGCTCAACCGCT-----T 3410
Qy 359 1eThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProA 379
Db 3411 TCATCTCTGAGAAAGCTTGATCTCAACACCGATGTGAGCAACATGAAGATTGTCCTGTG 3470
Qy 379 spValPheAlaGluAsnMellyb----- 386
Db 3471 ATGACTTTGTCAAAGCTGTGAAGGTACGTGGATAATGACTTTTTTTGTGACCGTGTC 3530
Qy 386 ----- 386
Db 3531 TTTGTCAACCGCTAACAACCTTCTTGAATCGGTCTCTTTGGTTTGAATTCGCTCGCG 3590
Qy 387 -----LeuArgGluAspThrHisIleuValAsnTyrIleProGlnGlySerI 403
Db 3591 CTTGACACAGGTCCGAGAGAAGACTCACACGCCGTGTCATATTGCCCCATCGGTTGCG 3650
Qy 403 1eAspSerLeuPheGluGlyThrTyrIleuValArgValAspGluLysHisArgArgT 423
Db 3651 TTGACGATCTCTGGCCTGATCGTACTACTTGGGAGAGATGACACGATGTGGCGTCGAC 3710
Qy 423 hrTyrAlaArgArgProThr 429
Db 3711 AGTACAAGCAGTCCCTTCT 3730

RESULT 10
US-09-925-388-1
; Sequence 1, Application US/09925388
; Patent No. 6586202
; GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: OJIMA, Kazuyuki
; APPLICANT: SETOBUCHI, Yutaka
; TITLE OF INVENTION: ISOPRENOID PRODUCTION
; FILE REFERENCE: ISOPRENOID PRODUCTION
; CURRENT APPLICATION NUMBER: US/09/925,388
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/306,595
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4775
; TYPE: DNA
; ORGANISM: Phaffia rhodozyma
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1239)..(1240)
; OTHER INFORMATION: EXPERIMENTAL
; NAME/KEY: exon
; LOCATION: (1305)..(1361)
; NAME/KEY: intron
; LOCATION: (1362)..(1504)
; NAME/KEY: exon
; LOCATION: (1505)..(1522)
; NAME/KEY: intron
```

```
; LOCATION: (1523)..(1699)
; NAME/KEY: exon
; LOCATION: (1700)..(1826)
; NAME/KEY: intron
; LOCATION: (1827)..(1920)
; NAME/KEY: exon
; LOCATION: (1921)..(2277)
; NAME/KEY: intron
; LOCATION: (2278)..(2351)
; NAME/KEY: exon
; LOCATION: (2352)..(2409)
; NAME/KEY: intron
; LOCATION: (2410)..(2497)
; NAME/KEY: exon
; LOCATION: (2498)..(2504)
; NAME/KEY: intron
; LOCATION: (2505)..(2586)
; NAME/KEY: exon
; LOCATION: (2587)..(2768)
; NAME/KEY: intron
; LOCATION: (2769)..(2851)
; NAME/KEY: exon
; LOCATION: (2852)..(2891)
; NAME/KEY: intron
; LOCATION: (2892)..(2985)
; NAME/KEY: exon
; LOCATION: (2986)..(3240)
; NAME/KEY: intron
; LOCATION: (3241)..(3325)
; NAME/KEY: exon
; LOCATION: (3326)..(3493)
; NAME/KEY: intron
; LOCATION: (3494)..(3601)
; NAME/KEY: exon
; LOCATION: (3602)..(3768)
; NAME/KEY: polyA site
; LOCATION: (4043)..(4044)
US-09-925-388-1

Alignment Scores:
Pred. No.: 1.36e-61 Length: 4775
Score: 585.00 Matches: 181
Percent Similarity: 37.15% Conservative: 82
Best Local Similarity: 25.56% Mismatches: 129
Query Match: 23.30% Indels: 318
DB: 4 Gaps: 15

US-10-622-516-2 (1-478) x US-09-925-388-1 (1-4775)
Qy 25 TyrPheProSerGln-----TyrValAspGlnAla----- 34
Db 1652 TATCTCCGACGCGAATAACAACACTGACCGCGATTCTCTCGATCAGGCCATCGCTCAC 1711
Qy 35 ---GluLeuGluLysTyrAspGlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGln 53
Db 1712 AAGGATCTCGAGGCTTTTGATGGGGTCTTCGGAAGTACACCATCGGTCTCGGCAAC 1771
Qy 54 AlaLysMetGlyPheCysThrAspArgGluAspIleAsnSerLeuCysMet-ThrVal-- 72
Db 1772 AACTTCATGCGCTTCACCGACGACACTGAGGACATCAACTCGTTCGCTTGAACGCTCAG 1831
Qy 72 ----- 72
Db 1832 TCTCTTCCGTTTCAGCAATCGACAGAAAAAGGCCCAAGCGCATCTCACTGACACCTTTC 1891
Qy 73 -----ValGlnAsnLeuMetGluArgAsnAsnLe 82
Db 1892 TCCGTTTGGCAATTCCATTGATTGTTAGCTGTTTCCGGTCTTCTAATCAAAAGTACAACT 1951
Qy 82 userTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLysSerly 102
Db 1952 TGATCCCAAGTCATCGGTGGAATTGATGTGGAACTGAGTCCATTCATTGACCAAGTCCAA 2011
```

QY 102 sSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIleGluG1 122
Db 2012 ATCTGTCAAGACAGATCCTTATGAGACTGTTCGAGTCCACGCGCAACAGATATGAGGG 2071
QY 122 yIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaValAsnTr 142
Db 2072 TATCGACTCCAGAAGATGCTCTGCTACGAGTTCACCGCGCTGTTCATGCGCTCACTG 2131
QY 142 pIleGluSerSerSerTrpAspGlyLeuArg----- 152
Db 2132 GATCGAGTCATCCTTGGAGCGAAG-AAATGCCATTGCTTCTGCGAGACATTGCCA 2190
QY 152 ----- 152
Db 2191 TCTACGCCGAGGGTGTGCGCGACCTGCCGAGGTGTGTGCTTGCGCCATCCTCATCG 2250
QY 152 ----- 152
Db 2251 GACCCGACGCTCCCGTGTCTTCGAGCGGTGAGTCCAAATCCGTCAATTTCTTCACGGCA 2310
QY 153 ----- 153
Db 2311 GCGGCTGAACAACCCCTTATCCGTCATCTTCATCAATGTAGCCGTCCACGAAACTTCAT 2370
QY 156 tGlnHisAlaTyrAspPheTyrLysProAspMetLeuSer----- 169
Db 2371 GACCAACGCTTGGAGACTCTACAGCCTAATCTTCTCGATGTCAATTTTGAAGTT 2430
QY 169 ----- 169
Db 2431 TGCGCTTGGAGAGTCTTACACTAATTCGGGGTGTCTGTATCCTTGAATCGTTGTTC 2490
QY 169 ----- 169
Db 2491 TTTATAGTGAATACGTTGCTGTGCGCACCTCCTATATTAAGTTTGTGATCAAAATATTGC 2550
QY 170 ----- 170
Db 2551 CATTGAATTAACCTTGAAACCTTCTCTCCAAATAG-CCCATGTGTGATGAGCCTCTCTC 2609
QY 179 rIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleH1 199
Db 2610 CGTCACTTCTACGTCAACGCCATGACAGCGCTATGAGCCTTACCGAACAAGTATGC 2669
QY 199 sAlaGlnTrpGlnLysGluGlyAsnAspLys----- 209
Db 2670 CAAGCGATTT-----GGAGAACCCCAAGACTAACGGTGTCAACCAACGACACCGCA 2720
QY 210 ----- 210
Db 2721 GGTGCGCGGTGTCAAGTGTGCTGTGCTGATTACCTTTGTCCACAGGTAAAGCGTCAATC 2780
QY 223 ----- 223
Db 2781 TTCTGTATTCTCCTTAATTTCAACCGATCAACGAGTTAATTCGTGTCAATATATTCT 2840
QY 226 ----- 226
Db 2841 TGTTGGAACAGTCTTACGAAAGCAGGTTGTCAAAAGGCCACGGCGACTTGTAAAGCAGT 2900
QY 236 ----- 236
Db 2901 CTTTGTTAACCTTAGCTTGACATAAAACTTTAGTTTCTGTACTATTATTAT 2960
QY 237 ----- 237
Db 2961 GCATCTCTTGAATCACCTTATAGTTGTACAAATGACTTCCGAAACAACCCCAAC----- 3015
QY 248 pLysAsnSerIleTyrSerGlyLeuGluAla--PheGlyAspValLysLeuGluAspTh 267
Db 3016 ----GACCCGGTTTGTGAGGTGCCAGCCGAGCCTTGCTACTTTGGACATGAAGAAAG 3071
QY 267 rTyrPheAspArgAspValGluLysAlaPheMetLysAlaSerSerGluLeuPheSerG1 287

Db 3072 TCTTTCAGACAGAAGATGTGAGAAATCTGTGATGTGCTGCCCTCAAGTCTTCTTCAACA 3131
QY 287 nLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyrThrSerSerVa 307
Db 3132 GCAGGTGAGCCTGGATGACCAACCGCTCCGACAGCTCGGAAACTGTATACACCGCTCTCT 3191
QY 307 lTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerProGlnGlnLeu----- 323
Db 3192 CTTCGGTGTCTCGCAAGTTTGTCTCTAATGTTCTGTGTCGAGCTCGTAAGTCTTGA 3251
QY 323 ----- 323
Db 3252 TCTCTATCCCAATCATCTTCTTATCAATTGAACGTAACTCTTTCTTTAATGCTGCG 3311
QY 324 ----- 324
Db 3312 TTTCTCTGAACAGGTGCGGCAAGCGCATTTGCTCTACGCCCTACGGATCTGAGCTGTG 3371
QY 339 lathrLeuTyrSerLeuLysValThrGlnAspAlathrProGlySerAlaLeuAspLysI 359
Db 3372 CTTCCTTCTATGCTCTTAAAGTCAAGAGCTCAACCGCT-----T 3410
QY 359 lathrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProA 379
Db 3411 TCATCTCTGAGAAGCTTGATCTCAACAACCGATTGAGCAACATGAAGATTGTCCTGTG 3470
QY 379 spValPheAlaGluAsnMetLys----- 386
Db 3471 ATGACTTGTCAAGCTCGAAGGTACGTTGATTAAGACTTTTTTTGTGACCGGTGTC 3530
QY 386 ----- 386
Db 3531 TTTGTCAACCGCTAACAACTTCTTGATCGTCTCTTTTGTGAAATTCGCTGCGG 3590
QY 387 ----- 387
Db 3591 CTTCGACACAGGTCCGAGAAGAGACTACACACGCCGTGCATATTGCGCCATCGGTGCG 3650
QY 403 leAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGluLysHisArgArgT 423
Db 3651 TTGACGATCTCTGGCGCTGATCTGTACTACTTGGAGAGATTGACAGCATGTGGCGTGCAC 3710
QY 423 hrTyrAlaArgArgProThr 429
Db 3711 AGTACAAGCAGGTCCCTTCT 3730

RESULT 11
US-08-998-416-730
; Sequence 730, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 730:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1476UP
;
US-08-998-416-730
```

```

Alignment Scores:
Pred. No.: 2.45e-56 Length: 635
Score: 530.50 Matches: 105
Percent Similarity: 62.44% Conservative: 23
Best Local Similarity: 51.22% Mismatches: 33
Query Match: 21.13% Indels: 44
DB: Gaps: 2
```

```

US-10-622-516-2 (1-478) x US-08-998-416-730 (1-635)
QY 29 GIntYrValaSpGlnAlaGluLeuGluYsTyrAspGlyValAspAlaGlyLysTyrThr 48
Db 20 CAGTGCCTGAACCAAGAGCATTTGAGGCGTATGACGCGCTGTCGCAAGGCAAGTACACT 79
QY 49 lLgLYleuGLyGlnAlaLysMetGlyPheCysThrAspArgGluAspIleAsnSerLeu 68
Db 80 ATCGGCTGGGCCAGACCAACATGAGCTTTGTGAACGACCGCGAGACATCTACTCGATG 139
QY 69 CysMetThrValValGlnAsnLeuMetGluArgAsnAsnLeuSerTyrAspCysIleGly 88
Db 140 TGTGTACCGCGCTGCTCGAATTGATGAAGAACTACGATATCAAGCCGGAAGCATCGGC 199
QY 89 ArgLeuGluValGlyThrGluThrIleIleAspLysSerLysSerValLysThrAsnLeu 108
Db 200 CGCCTCGAGGTGGGTACGGAGACGTTGCTTGACAAAGTCGAAGTCCGTGAAGTCTATT 259
QY 109 MetGlnLeuPheGluGluSerGlyAsnThrAspIleGluGlyIleAspThrThrAsnAla 128
Db 260 ATGCAGTTGTTCCGGCAG-----AACACCGACTTGGAGGGTGTGATACCGTGAACGCC 313
QY 129 CysTyrGlyGlyThrAlaAlaValPheAsnAlaValAsnTrpIleGluSerSerSerTrp 148
Db 314 TGCTATGGCGGTAATAACGCCGTTGTTAACTCTTGAAGTGAAGTGAAGTCCAGTTCGTGG 373
QY 149 AspGly----- 150
Db 374 GACGGTCGTGACCAATCGTTGTTGTGTGATCGCAATCTACGACAAGGGTGCCGCC 433
QY 150 ----- 150
Db 434 CGGCCCACTGGCGGTGCGGAACGTGCTCTCTGATCGGTCCAGACGCCCCCATTTGTC 493
QY 151 -----LeuArgGlyThrHisMetGlnHisAlaTyrAspPheTyrLysProAspMet 167
Db 494 TTGACTCTGTGCGGTGCTCGTACATGAGACACGCTCTACGACTTCTCAAGCCCTGACTTC 553
QY 168 LeuSerGluTyrProIleValAspGlyLysLeuSerIleGlnCysTyrLeuSer-AlaIle 187
```

```

Db 554 CGCAGTGAAGTATCCATACGTGACGCGCCACTTCTCACTAACATGCTACGTCAAGGCCGT 613
QY 187 uAspArgCysTyr 191
Db 614 CGACCAAGGCTTAC 626
```

```

RESULT 12
US-09-833-381-1648
; Sequence 1648, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186e1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1648
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(472)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1648
```

```

Alignment Scores:
Pred. No.: 1.13e-39 Length: 472
Score: 395.00 Matches: 86
Percent Similarity: 85.85% Conservative: 5
Best Local Similarity: 81.13% Mismatches: 12
Query Match: 15.73% Indels: 4
DB: Gaps: 0
```

```

US-10-622-516-2 (1-478) x US-09-833-381-1648 (1-472)
QY 376 ValAlaProAspValPheAlaGluAsnMetLys-LeuArgGluAspThrHisIleuVa 395
Db 8 GTGGCACCAGATGCTTGTCTGAAAACATGAATGCTCAGAGAGACACACATCACTTAGC 67
QY 395 lAsnTyrIleProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValAr 415
Db 68 CAATAATATTCCTCCATGT-TCATATNACTCACTCTTGAAGAAAGTGTATCTGTGTCAG 126
QY 415 gValAspGluLysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLe 435
Db 127 AGTGATGAATAAAACACAGAAGAGACTTACGCCCGCGCCCTTCAACAATGACACAGATT 186
QY 435 uAspGluGlyValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAl 455
Db 187 GGATGAAGGAATGNNCTCGTGCATAGTAAACACAGCAAGACATATTCCAAGCCCTGC 246
QY 455 aLysLysValProArgLeuProAlaThrAlaAlaGluProGluAla-AlaValIleSer- 474
Db 247 TAAGAAAGTGCAAGACTCCCTGCAACCTCGCGCGAATCTGAATCACNCTGTATCATCAT 306
QY 475 AsnGlyLysHis 478
Db 307 AACGGGAGACAC 318
```

```

RESULT 13
US-09-107-532A-1715
; Sequence 1715, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
```



```

; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1715:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1167 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1167
; SEQUENCE DESCRIPTION: SEQ ID NO: 1715:
US-09-107-532A-1715

Alignment Scores:
Pred. No.: 5.16e-29 Length: 1167
Score: 315.00 Matches: 104
Percent Similarity: 40.10% Conservative: 64
Best Local Similarity: 24.82% Mismatches: 135
Query Match: 12.54% Indels: 116
DB: 4 Gaps: 13

US-10-622-516-2 (1-478) x US-09-107-532A-1715 (1-1167)
QY 17 ValGlyIleValAlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeu 36
Db ::::::::::::::::::::
19 ATAGGCAATGCATCGTCTTCCCTTTTATTCTCAATTATATTAGATATGACAGATTG 78
QY 37 GluLysTyrAspGlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMet 56
Db ::::::::::::::::::::
79 GCAGAAAGCCGTGGGATGATCTGCAAAATACCATATAGGATTGGTCAAGACCAATG 138
QY 57 GlyPheCysThrAspArgGluAspIleAsnSerLeuCysMetThrValAlaGlnAsnLeu 76
Db 139 GCAGTCAATCGTCAAAATGAAGCATCATTTACACTAGAGCAAAATGCTGCCAGCAAAATC 198
QY 77 MetGluArgAsnAsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThr 96
Db 199 GTAACAGAAAAAGAC-----CGTAGACTAATCGACATGTCATCGTCCGAACAGAAATCC 252
QY 97 IleIleAspLysSerLysSerValLysThrAsnLeu-----MetGlnLeu 111
::::::::::::::::::
```

```

Db 253 GGGATTGATCATTCAAAAGCGAGTGGGTAATTATCCACCATTACTGAAAATCCAACTCT 312
QY 112 PheGluGluSerGlyAsnThrAspIleGluGlyIleAspThrThrAsnAlaCysTyrGly 131
Db ::::::::::::::::::::
313 TTTCCTCGTCTCT-----TTTGAAGTAAAAAGAGCTTGCTACGGT 351
QY 132 GlyThrAlaAlaValPheAsnAlaValAsnTrpIleGlu----- 144
Db ::::::::::::::::::::
352 GGCACCCGACGCTTGACATGCGCAAGAAAGATATGTCAAAAATCATCCAGAACGAAAGTA 411
QY 145 -----SerSerSerTrpAspGlyLeuArg 152
Db 412 CTAGTCATAGCAAGTATATTGCTCTGTTACGGCTTGCGCAAGCGGTGTAAGTCAGCAA 471
QY 153 GlyThrHis----- 155
Db 472 GGTGTCGGTCTGTGGCATGATGATCACTCAAAACCCGCTATTATTATCGATTGAAGAC 531
QY 156 -----MetGlnHisAlaTyrAspPheTyrLysProAspMetLeuSerGlu 170
Db 532 GACAGCGTATTCTGACAGAAAGATATCTATGATTCTGCGCTCCAGAT--TATAGCGAA 588
QY 171 TyrProIleValAspGlyLysLeuSerIleGlnCysTyrLeuSerAlaLeuAspArgCys 190
Db ::::::::::::::::::::
589 TTTCCTGTGTGATGATGCTTCTTATCTAATTCATGTAATATCGAATCATTCCAAAAAGTT 648
QY 191 TyrSerValTyrCysLysLysIleHisAlaGlnTrpGlnLysGluGlyAsnAspLysAsp 210
Db 649 -----TGAATCGACATAAAGATTGTCCGGT 675
QY 211 PheThrLeuAsnAspPheGlyPheMetIlePheHisSerProTyrCysLysLeuValGln 230
Db 676 CGAGGACTCGAAGATTATCAGCGCATTTGCTTCCATATTCCGTATACTAAGATGGGAAA 735
QY 231 LysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsnAspGlnAsnArgAspLysAsn 250
Db 736 AAGCATTCGAAAGCGTATTA-----GACCAAAACAGACGAAGAAAT 777
QY 251 SerIleTyrSerGlyLeuGluAlaPheGlyAspValLysLeuGluAspThrTyrPheAsp 270
Db 777 ----- 777
QY 271 ArgAspValGluLysAlaPheMetLysAlaSerSerGluLeuPheSerGlnLysThrLys 290
Db 778 -----CAGGAACGCTTATGCGTCGCTATGAA 804
QY 291 AlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyrThrSerSerValTyrGlySer 310
Db 805 GAAAGCATCCGTTACAGCCGACGCAATCGGTAATCTTACACTGTTCAATTATACCTGGGG 864
QY 311 LeuAlaSerValLeuAlaGlnTyrSerProGlnGlnLeuAlaGlyLysArgIleGlyVal 330
Db 865 CTAACCTCTCTACTG--GAAAATTGAAATCACTACAGCCAGAGATCGCATCGTCTA 921
QY 331 PheSerTyrGlySerGlyLeuAlaAlaThrLeuTyrSerLeuLysValThrGlnAspAla 350
Db 922 TTCAGCTACGGTTCTGGTCAGTAAAGTGAATTTTTCACAGTTATCTGAAAGAAACTAT 981
QY 351 ThrProGlySerAlaLeuAspLysIleThrAlaSerLeuCysAspLeuLysSerArgLeu 370
Db 982 CAA-----GAATATCTCTTCGCACAA-----TCTCATCAAGAGATGCTA 1020
QY 371 AspSerArgThrGlyValAlaProAsp-----ValPheAlaGluAsnMet 385
Db 1021 GATTCTGACACCGGATCACTGTGTGACGAATATGAACGATTTTTCAGTGAGACGCTT 1077

RESULT 14
US-08-956-171E-364
; Sequence 364, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunesh
; Gil H. Choi
```

[illegible]

QY	117	AsnThrAspIleGluGlyIle-----AspThrThrAsnAlaCys	129
		-----GGTATTCAACCTTTTGCACGCGCTTGGAATGAAGAAGCTTGT	4157
Db	4113	-----	-----
QY	130	TyrGlyGlyThrAlaAlaValPheAsnAlaValAsnTrpIleGlu-----	144
		-----	-----
Db	4158	TATGCTGCACAACCAACCAATTCAATTAGCTTAAGATTATTAGCAACTAGACCGAATGAA	4217
QY	145	-----SerSerSerTrpAspGly	150
Db	4218	AAAGTATTAGTTATTGCTACAGATACAGACGTTATGATTGAATTCAGCGCGAGCCA	4277
QY	151	LeuArgGly-----	153
Db	4278	ACACACAGTGTCTGGCGCAGCTTGCAGTGTATTGCACATTAATCCAAAGCATTTTGGCATTA	4337
QY	154	-----ThrHisMetGlnHisAlaTyrAspPheTyrLysProAspMetLeu	168
Db	4338	AATGAAGATGCTGTGCTTACACTGAAGACGTTTATGATTTCGGCGTCCAACTGGA--	4394
QY	169	SerGluTyrProIleValAspGlyLysLeuSerIleGlnCysTyrLeuSerAlaLeuAsp	188
Db	4395	CATAAATATCCATTAGTTCATGTGTCATTTCTTAAGANGCTTATATCCGCTCATTTCCA	4454
QY	189	ArgCysTyrSerValTyrCysLysLysIleHisAlaGlnTrpGlnLysGluGlyAsnAsp	208
Db	4455	CAAAGCTGCAATGAATACGCCAAA-----CGTCAAGGTTAG--	4490
QY	209	LysAspPheThrLeuAsnAspPheGlyPheMetIlePheHisSerProTyrCysLysLeu	228
Db	4491	-----TCGCTAGCTGACTTCGCATCTCTATGCTTCCATGTTCATTACAAAATG	4541
QY	229	ValGlnLysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsnAspGlnAsnArgAsp	248
Db	4542	GCTAAAGGCAATTAGAGTCAATCATTTGATTAACGCTGATGAACAACTCAAGAGCGT--	4598
QY	249	LysAsnSerIleTyrSerGlyLeuGluAlaPheGlyAspValLysLeuGluAspThrTyr	268
Db	4599	-----TTACGTTCAAGATATGAAGATGCTGTAGAT-----	4628
QY	269	PheAspArgAspValGlnLysAlaPheMetLysAlaSerSerGlnLeuPheSerGlnLys	288
Db	4629	TATAACCGTTATGTC-----	4643
QY	289	ThrLysAlaSerLeuLeuValSerAsnGlnAsnGlnAsnMetTyrThrSerSerValTyr	308
Db	4644	-----GGTAATATTATTACTGATCATTTATAT	4670
QY	309	GlySerLeuAlaSerValLeuAlaGlnTyrSerProGlnGlnLeuAlaGlyLysArgIle	328
Db	4671	TTAAGCCTAATATCATTTACTTGAAATCGAGATTTACAA-----GCTGGTGAACAATC	4724
QY	329	GlyValPheSerTyrGlySerGlyLeuAlaAlaThrLeuTyrSerLeuLysValThrGln	348
Db	4725	GGTTTATTCAATTATGGCTCAGGTTCAAGTTGGTGAATTTTATAGTGCACATTAGTTGAA	4784
QY	349	AspAlaThrProGlySerAlaLeuAspLysIleThrAlaSerLeuCysAspLeuLysSer	368
Db	4785	GGCTACAA-----GATCATTTAGATCAAGCTGCACAT-----AAAGCA	4823
QY	369	ArgLeuAspSerArgThrGlyValAlaProAspValPheAlaGluAsnMetLysLeuArg	388
Db	4824	TTATTAATAATAACCGTACTGAAGTATCTGTTGATGCATATGAACAATCTTCAAAAGTTT	4883
QY	389	GluAsp	390
Db	4884	GATGAC	4889

RESULT 15

US-09-134-001C-1458

/ Sequence 1458, Application US/09134001C

/ Patent No. 6380370

/ GENERAL INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2004, 13:02:01 ; Search time 49 Seconds
(without alignments)
938.358 Million cell updates/sec

Title: US-10-622-516-2

Perfect score: 2511

Sequence: 1 MPGSLPLNAEACWPKDVGIV.....PRLPATAEPEAAVISNGEH 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2480	98.8	520	2 S45497	hydroxymethylgluta
2	2439	97.1	520	2 S27197	hydroxymethylgluta
3	2362	94.1	520	2 S12736	hydroxymethylgluta
4	2352	93.7	520	2 A25332	hydroxymethylgluta
5	2095	83.4	522	2 S13887	hydroxymethylgluta
6	1513	60.3	508	2 S71623	hydroxymethylgluta
7	1494	59.5	508	2 A35865	hydroxymethylgluta
8	1493	59.5	471	2 B55729	hydroxymethylgluta
9	1352.5	53.9	453	2 S38986	hydroxymethylgluta
10	1272.5	50.7	455	2 A53565	hydroxymethylgluta
11	1009.5	40.2	474	2 T09688	hydroxymethylgluta
12	983.5	39.2	454	2 T49718	probable hydroxyme
13	980	39.0	461	2 T09341	hydroxymethylgluta
14	950.5	37.9	491	2 S58202	hydroxymethylgluta
15	900	35.8	447	2 S61875	hydroxymethylgluta
16	705	28.1	462	2 T25726	hypothetical prote
17	338.5	13.5	98	2 S22432	hydroxymethylgluta
18	312.5	12.4	445	2 A84314	3-hydroxy-3-methyl
19	301.5	12.0	388	2 C90059	3-hydroxy-3-methyl
20	299.5	11.9	384	2 D86821	hydroxymethylgluta
21	272.5	10.9	388	2 A81614	hydroxy-3-methylgl
22	267.5	10.7	388	2 A61251	hydroxy-3-methylgl
23	244	9.7	398	2 C95201	hydroxymethylgluta
24	241	9.6	398	2 A98068	hypothetical prote
25	211	8.4	412	2 AG0177	probable hydroxyme
26	204	8.1	407	2 B70185	probable hydroxyme
27	158	6.3	42	2 S13490	hydroxymethylgluta
28	147	5.9	345	2 E69678	involved in polyke
29	145.5	5.8	349	2 F69205	3-hydroxy-3-methyl

30	144	5.7	350	2 F71113	probable acyl carr
31	138.5	5.5	350	2 T45082	acyl carrier prote
32	126.5	5.0	350	2 E75047	acyl carrier prote
33	123.5	4.9	716	2 J00275	beta-galactosidase
34	122	4.9	345	2 A64493	acyl carrier prote
35	114.5	4.6	458	2 C90513	conserved hypothet
36	114	4.5	1486	2 A10906	glutamate synthase
37	112	4.5	549	2 C89930	alpha-D-1,4-glucos
38	108.5	4.3	1619	2 T18499	hypothetical prote
39	108	4.3	464	2 A87675	carboxyl-terminal
40	107.5	4.3	776	2 H71688	DNA topoisomerase
41	106.5	4.2	766	2 T03218	armadillo-like pro
42	106.5	4.2	1517	1 F65112	glutamate synthase
43	106	4.2	367	2 T20271	hypothetical prote
44	105.5	4.2	1144	2 A54810	TMV resistance pro
45	105.5	4.2	1517	2 C91140	glutamate synthase

ALIGNMENTS

RESULT 1					
S45497					
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5), cytosolic, adrenal isoform - human					
N;Alternate names: 3-hydroxy-3-methylglutaryl coenzyme A synthase					
C;Species: Homo sapiens (man)					
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-May-2000					
C;Accession: S45497; I39355					
R;Rokosz, L.L.; Boulton, D.A.; Butkiewicz, E.A.; Sanyal, G.; Cueto, M.A.; Lachance, P.A.,					
Arch. Biochem. Biophys. 312, 1-13, 1994					
A;Title: Human cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase: expression, p					
A;Reference number: I39355; MUID:94304197; PMID:7913309					
A;Accession: S45497					
A;Molecule type: mRNA					
A;Residues: 1-520 <ROK>					
A;Cross-references: EMBL:L25798; NID:g410027; PIDN:AAA62411.1; PID:g410028					
A;Experimental source: fetal adrenal					
C;Function:					
A;Description: catalyzes the condensation of acetyl coenzyme A (Ac-CoA) with acetoacetyl					
tase					
C;Superfamily: hydroxymethylglutaryl-CoA synthase					
C;Keywords: carbon-carbon lyase; cholesterol biosynthesis; coenzyme A; oxo-acid-lyase					
F;129/Active site: Cys (covalent substrate-binding) #status predicted					
Query Match 98.8%; Score 2480; DB 2; Length 520;					
Best Local Similarity 91.9%; Pred. No. 4.4e-177;					
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;					
Qy	1	MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQAKGCT	60		
Db	1	MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKYYTIGLGQAKGCT	60		
Qy	61	DREDINSLCMTVVQNLMERNLSYDCIGRLVGTETIIIDKSQSVKTNLMQLFEESGNTDI	120		
Db	61	DREDINSLCMTVVQNLMERNLSYDCIGRLVGTETIIIDKSQSVKTNLMQLFEESGNTDI	120		
Qy	121	EGIDTTNACYGTAAVFNAVWMISSSWD-----	149		
Db	121	EGIDTTNACYGTAAVFNAVWMISSSWDGRYALVVAGDIAVYATGNARPTGSGAVALL	180		
Qy	150	-----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVCKKI	198		
Db	181	IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVCKKI	240		
Qy	199	HAQWQKEGNDKDFTLNDGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGIEA	258		
Db	241	HAQWQKEGNDKDFTLNDGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGIEA	300		
Qy	259	FGDVKLEDTYFDRDVEKAFMKASSELFSQKTAKSLVSNQNGMYTSSVYGSGLAVLAQY	318		
Db	301	FGDVKLEDTYFDRDVEKAFMKASSELFSQKTAKSLVSNQNGMYTSSVYGSGLAVLAQY	360		
Qy	319	SPQQLAGKRIGVFSYSGIAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP	378		

```
Db      361 SPQOLAGKRIGVFSYSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY      379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEG 438
Db      421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEG 480
QY      439 VGLVHSNIATEHIIPSPAKKVPRLPATAAEPAAVISNGEH 478
Db      481 VGLVHSNIATEHIIPSPAKKVPRLPATAAEPAAVISNGEH 520
```

RESULT 2

```
S27197
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5), cytosolic, fibroblast isoform - human
C:Species: Homo sapiens (man)
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-May-2000
C:Accession: S27197; S21590
R:Russ, A.P.; Ruzicka, V.; Maerz, W.; Appelhans, H.; Gross, W.
Biochim. Biophys. Acta 1132, 329-331, 1992
A:Title: Amplification and direct sequencing of a cDNA encoding human cytosolic 3-hydrox-
A:Reference number: S27197; MUID:93041939; PMID:1358203
A:Accession: S27197
A:Molecule type: mRNA
A:Residues: 1-520 <RUS>
A:Cross-references: EMBL:X66435; NID:g30008; PIDN:CAA47061.1; PID:g30009
A:Experimental source: fibroblast
C:Function:
A:Description: catalyzes the condensation of acetyl coenzyme A (Ac-CoA) with acetoacetyl-
ase
C:Superfamily: hydroxymethylglutaryl-CoA synthase
C:Keywords: carbon-carbon lyase; cholesterol biosynthesis; coenzyme A; oxo-acid-lyase
F:129/Active site: Cys (covalent substrate-binding) #status predicted
```

```
Query Match      97.1%; Score 2439; DB 2; Length 520;
Best Local Similarity 90.9%; Pred. No. 5e-174;
Matches 471; Conservative 1; Mismatches 4; Indels 42; Gaps 1;
```

```
QY      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKTYTIGLQAKMGFCT 60
Db      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKTYTIGLQAKMGFCT 60
QY      61 DREDINSLCMTVVQNLMEERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db      61 DREDINSLCMTVVQNLMEERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY      121 EGIDTTNACYGGTAAVFNAVNMISSSSWD----- 149
Db      121 EGIDTTNACYGGTAAVFNAVNMISSSSWDGRYALVAGDIAIYASGNARPTGCVAVALL 180
QY      150 -----GLRGTHMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKKI 198
Db      181 IGPNAVIFDRGLRGTHMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKKI 240
QY      199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db      241 HAQWQKEANDNDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY      259 FGDVKLIEDTYFDRDVEKAFMKASSELFSQKTASLVSNOQNGMYTSSVYGLASVLAQY 318
Db      301 FGDVKLIEDTYFDRDVEKAFMKASSELFSQKTASLVSNOQNGMYTSSVYGLASVLAQY 360
QY      319 SPQOLAGKRIGVFSYSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db      361 SPQOLAGKRIGVFSYSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY      379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEG 438
Db      421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEG 480
QY      439 VGLVHSNIATEHIIPSPAKKVPRLPATAAEPAAVISNG 476
Db      481 VGLVHSNIATEHIIPSPAKKVPRLPATAAEPAAVISNG 518
```

RESULT 3

```
S12736
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5), cytosolic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-May-2000
C:Accession: S12736
R:Ayte, J.; Gil-Gomez, G.; Hegardt, F.G.
Nucleic Acids Res. 18, 3642, 1990
A:Title: Nucleotide sequence of a rat liver cDNA encoding the cytosolic 3-hydroxy-3-meth-
A:Reference number: S12736; MUID:90301491; PMID:1972979
A:Accession: S12736
A:Molecule type: mRNA
A:Residues: 1-520 <AYT>
A:Cross-references: EMBL:X52625; NID:g55946; PIDN:CAA36852.1; PID:g55947
C:Superfamily: hydroxymethylglutaryl-CoA synthase
C:Keywords: carbon-carbon lyase; coenzyme A; cytosol; oxo-acid-lyase
```

```
Query Match      94.1%; Score 2362; DB 2; Length 520;
Best Local Similarity 87.5%; Pred. No. 2.8e-168;
Matches 455; Conservative 11; Mismatches 12; Indels 42; Gaps 1;
```

```
QY      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKTYTIGLQAKMGFCT 60
Db      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAKTYTIGLQAKMGFCT 60
QY      61 DREDINSLCMTVVQNLMEERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db      61 DREDINSLCMTVVQNLMEERNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY      121 EGIDTTNACYGGTAAVFNAVNMISSSSWD----- 149
Db      121 EGIDTTNACYGGTAAVFNAVNMISSSSWDGRYALVAGDIAIYASGNARPTGCVAVALL 180
QY      150 -----GLRGTHMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKKI 198
Db      181 IGPNAVIFDRGLRGTHMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVCKKI 240
QY      199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db      241 RAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY      259 FGDVKLIEDTYFDRDVEKAFMKASSELFSQKTASLVSNOQNGMYTSSVYGLASVLAQY 318
Db      301 FGDVKLIEDTYFDRDVEKAFMKASSELFSQKTASLVSNOQNGMYTSSVYGLASVLAQY 360
QY      319 SPQOLAGKRIGVFSYSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db      361 SPQOLAGKRIGVFSYSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY      379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEG 438
Db      421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEG 480
QY      439 VGLVHSNIATEHIIPSPAKKVPRLPATAAEPAAVISNGEH 478
Db      481 VGLVHSNIATEHIIPSPAKKVPRLPATSGEPESAVISNGEH 520
```

RESULT 4

```
A25332
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Chinese hamster
C:Species: Cricetus griseus (Chinese hamster)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 05-May-2000
C:Accession: A25332
R:Gil, G.; Goldstein, J.L.; Slaughter, C.A.; Brown, M.S.
J. Biol. Chem. 261, 3710-3716, 1986
A:Title: Cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase from the hamster.
A:Reference number: A25332; MUID:86140166; PMID:2869035
A:Accession: A25332
A:Molecule type: mRNA
A:Residues: 1-520 <GIL>
```


A;Cross-references: GB:L00334; NID:g191380; PIDN:AAA37076.1; PID:G387072
A;Note: the source is Chinese hamster ovary cells
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; cholesterol biosynthesis; coenzyme A; endoplasmic retic

Query Match 93.7%; Score 2352; DB 2; Length 520;
Best Local Similarity 87.5%; Pred. No. 1.5e-167;
Matches 455; Conservative 9; Mismatches 14; Indels 42; Gaps 1;

QY 1 MFGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTIGLGQAKMGFCT 60
Db 1 MFGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTIGLGQAKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNLSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNLSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY 121 EGIDDTNACYGCTAAVFNAVNWIESSWD----- 149
Db 121 EGIDDTNACYGCTAAVFNAVNWIESSWDGRYALVVA GDIAIYATGNARPTGAGAVALL 180
QY 150 -----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLISQCYLSALDRCYSVYCKKI 198
Db 181 IGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPIVDGKLISQCYLSALDRCYSVYRKKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
Db 241 RAQWQKEGNDNDFTLNDFGFMISHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTASLLVSNQNGNMYTSSVYGSILASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASSELFNQKTASLLVSNQNGNMYTSSVYGSILASVLAQY 360
QY 319 SPQQLAGKRIGVFSYSGSLAATLYSLKVTODATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db 361 SPQQLAGKRIGVFSYSGSLAATLYSLKVTODATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVFAENMKLRDETHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDTLDG 438
Db 421 DVFAENMKLRDETHHLVNYIPQCSIDSLFEGTWYLVRVDEKHRRTYARRPSTNDHNLGDG 480
QY 439 VGLVHSNIATEHIPSPAKKVPRLPATAEPEAAVISNGEH 478
Db 481 VGLVHSNTATEHIPSPAKKVPRLPATAESESAVISNGEH 520

RESULT 5
S13887
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - chicken
C;Species: Gallus gallus (chicken)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-May-2000
C;Accession: S13887
R;Kattar-Coolley, P.A.; Wang, H.H.L.; Mende-Mueller, L.M.; Miziorko, H.M.
Arch. Biochem. Biophys. 283, 523-529, 1990
A;Title: Avian liver 3-hydroxy-3-methylglutaryl-CoA synthase: distinct genes encode the
A;Reference number: S13887; MUID:91112772; PMID:1980405
A;Accession: S13887
A;Molecule type: mRNA
A;Residues: 1-522 <KAT>
A;Cross-references: EMBL:M60657
A;Note: the authors translated the codon CTG for residue 280 as Ile
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 83.4%; Score 2095; DB 2; Length 522;
Best Local Similarity 76.4%; Pred. No. 2.3e-148;
Matches 399; Conservative 37; Mismatches 42; Indels 44; Gaps 3;

QY 1 MFGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGDAGKYTIGLGQAKMGFCT 60
Db 1 MFGSLPVNTESCWPKDVGIVALEIYFPSQYVDQTELEKYDGDAGKYTIGLGQSKMGFCS 60
QY 61 DREDINSLCMTVVQNLMERNLSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

QY 61 DREDINSLCMTVVQNLMERNLSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDV 120
QY 121 EGIDDTNACYGCTAAVFNAVNWIESSWD----- 149
Db 121 EGIDDTNACYGCTAALFNAINWIESSWDGRYALVVA GDIAVYATGNARPTGAGAVAML 180
QY 150 -----GLRGTHMQHAYDFYKPDMLSEYPIVDGKLISQCYLSALDRCYSVYCKKI 198
Db 181 VGSNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLISQCYLSALDRCYSVYRKKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRD-KNSIYSGLE 257
Db 241 HAQWQKEGTRGFTLNDFGFMIFHSPYCKLVQKSVARLLINDFLSDQNAETANGVFSGLE 300
QY 258 AFGDVKLEDTYFDRDVEKAFMKASSELFSQKTASLLVSNQNGNMYTSSVYGSILASVLAQ 317
Db 301 AFRDVKLEDTYFDRDVEKAFMKASAELEFNQKTASLLVSNQNGNMYTPSVYGCILASLLAQ 360
QY 318 YSPQQLAGKRIGVFSYSGSLAATLYSLKVTODATPGSALDKITASLCDLKSRLDSRTGVA 377
Db 361 YSPEHLAGQRISFEFSYSGFAATLYSIRVTODATPGSALDKITASLSDLKARLDSRKCIA 420
QY 378 PDVFAENMKLRDETHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDTLD 437
Db 421 PDVFAENMKIRQETHHLVNYIPQCSVEDLFEGTWYLVRVDEKHRRTYARRPVGDGPLEA 480
QY 438 VGLVHSNIATEHIPSPAKKVPRLPAT-AEPEAAVISNGEH 478
Db 481 GVEVHPGIVHEHIPSPAKKVPRIPATTESEGTVAISNGVH 522

RESULT 6
S71623
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) precursor, mitochondrial - human
C;Species: Homo sapiens (man)
C;Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 05-May-2000
C;Accession: S71623; A55729; S51103
R;Mascaro, C.; Buesa, C.; Ortiz, J.A.; Haro, D.; Hegardt, F.G.
Arch. Biochem. Biophys. 317, 385-390, 1995
A;Title: Molecular cloning and tissue expression of human mitochondrial 3-hydroxy-3-methy
A;Reference number: S71623; MUID:95200282; PMID:7893153
A;Accession: S71623
A;Molecule type: mRNA
A;Residues: 1-508 <MAS>
A;Cross-references: EMBL:X83618; NID:g619876; PIDN:CAA58593.1; PID:g619877
A;Experimental source: liver
R;Boukaftane, Y.; Duncan, A.; Wang, S.; Labuda, D.; Robert, M.F.; Sarrazin, J.; Schappert
Genomics 23, 552-559, 1994
A;Title: Human mitochondrial HMG CoA synthase: liver cDNA and partial genomic cloning, cf
A;Reference number: A55729; MUID:95154824; PMID:7851882
A;Accession: A55729
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 38-508 <BOU>
A;Cross-references: GB:U12788; GB:U12789
C;Genetics:
A;Gene: GDB:HMGC52
A;Cross-references: GDB:342071; OMIM:600234
A;Map position: 1p13-1p12
A;Genome: nuclear
C;Function:
A;Description: catalyses the condensation of acetyl-CoA and acetoacetyl-CoA to hydroxymet
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; mitochondrion; oxo-acid-lyase
F;1-37/Domain: transit peptide (mitochondrion) #status predicted <SIG>
F;38-508/Product: hydroxymethylglutaryl-CoA synthase #status predicted <MAT>
F;166/Active site: Cys (covalent substrate-binding) #status predicted

Query Match 60.3%; Score 1513; DB 2; Length 508;
Best Local Similarity 60.4%; Pred. No. 5.9e-105;
Matches 282; Conservative 69; Mismatches 74; Indels 42; Gaps 1;

QY 4 SLPNAEACWPKDVGIVALEIYFPPSQYVDQAELEKYDGVDAKYYTIGLGOAKMGFCTDRE 63
Db 41 AVFLAKTDTWPKDVGILALEVYFPAQYVDQDTLEKYNVNEAGKYTVGLGQTRMGFCSVQE 100
QY 64 DINSICMTVVQNLMERNNLSDYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEG1 123
Db 101 DINSICLTVQRLMERIQLPMDSVGRLEVGTEITIIDKSKAVKTVLMELFQDSGNTDIEG1 160
QY 124 DTNACYGGTAAVFNAVNMIESSWD----- 149
Db 161 DTNACYGGTASLFNNAANMMESSWDGRYAMVVCGDIAVYPSGNARPTGAGAVAMLIGP 220
QY 150 -----GLRGTHMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQ 201
Db 221 KAPLALERGLRGTHMENAYDFYKPNLASEYPIVDGKLSIQCYLRALDRCYTSYRKTIQNO 280
QY 202 WQEGNDKFTLNDFGFEMIHFSPYCKLVOKSLARMLLNDFLNDQNRDKNSIYSGLEAFGD 261
Db 281 WKQAGSDRPFLLDDQYMIFHTPFCKMVOKSLARLMFNDFLSASSDTQTSLYKGLEAFGG 340
QY 262 VKLEDYFDRDVEKAFMKASSELFSQKTKASLLVSNONGNMYTSSVYGSILASVLAQYSPQ 321
Db 341 LKLEDYTNKDLKALKLAKASQDMFDKTKASLYLSTNGNMYTSSLYGCLASLSHSAQ 400
QY 322 QLAGKRIGVFSYSGSLAATLYSLKVTODATPGSALDKITASLCDLKSRLDSRTGVAPDVF 381
Db 401 ELAGSRIGAFSYSGSLAASFSSFRVSDAAPGSPDLKLVSTSDLPKRLASRKCVSPEEF 460
QY 382 AENMKLREDTHLNVNIPQGSIDSLFEGTWYLVRVDEKGRRTYARRP 428
Db 461 TEIMNQREQFYHKVNFSPPGDTNSLFPGTWYLERVDEQHRKRYARRP 507

RESULT 7

A35865 hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) precursor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 05-May-2000
C/Accession: A35865, S32477
R/Ayte, J.; Gil-Gomez, G.; Haro, D.; Marrero, P.F.; Hegardt, F.G.
Proc. Natl. Acad. Sci. U.S.A. 87, 3874-3878, 1990
A/Title: Rat mitochondrial and cytosolic 3-hydroxy-3-methylglutaryl-CoA synthases are en
A/Reference number: A35865, MUID:90251660, PMID:1971108
A/Accession: A35865
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-508 <AYT>
A/Cross-references: GB:M33648, NID:g204618, PIDN:AAA41336.1, PID:g204619
Eur. J. Biochem. 213, 773-779, 1993
A/Title: The rat mitochondrial 3-hydroxy-3-methylglutaryl-coenzyme-A-synthase gene conta
A/Reference number: S32477, MUID:93238765, PMID:8097464
A/Accession: S32477
A/Molecule type: DNA
A/Residues: 1-35 <GIL>
A/Cross-references: EMBL:M63800, NID:g294590
A/Experimental source: strain Sprague-Dawley; tissue liver
A/Note: the sequence shown follows the authors' translation at position 35
C/Genetics:
A/Genome: nuclear
C/Superfamily: hydroxymethylglutaryl-CoA synthase
C/Keywords: carbon-carbon lyase; cholesterol biosynthesis; coenzyme A; mitochondrion; ox
F;1-37/Domain: transit peptide (mitochondrion) #status predicted <Sig>
F;38-508/Product: hydroxymethylglutaryl-CoA synthase #status predicted <MAT>
F;166/Active site: Cys (covalent substrate-binding) #status predicted

Query Match 59.5%; Score 1494; DB 2; Length 508;
Best local similarity 59.1%; Pred. No. 1.5e-103;
Matches 275; Conservative 76; Mismatches 72; Indels 42; Gaps 1;

QY 6 PLNAEACWPKDVGIVALEIYFPPSQYVDQAELEKYDGVDAKYYTIGLGOAKMGFCTDREDI 65
Db 43 PLAKTDTWPKDVGILALEVYFPAQYVDQDTLEKYNVNEAGKYTVGLGQTRMGFCSVQEDI 102

QY 66 NSLCMTVVQNLMERNNLSDYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEG1DT 125
Db 103 NSLCITVQRLMERITKLPMDAVGRLEVGTEITIIDKSKAVKTVLMELFQDSGNTDIEG1DT 162
QY 126 TNACYGGTAAVFNAVNMIESSWD----- 149
Db 163 TNACYGGTASLFNNAANMMESSYWDGRYALVVCGDIAVYPSGNARPTGAGAVAMLIGPKA 222
QY 150 -----GLRGTHMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQ 203
Db 223 PLVLEQGLRGTHMENAYDFYKPNLASEYPIVDGKLSIQCYLRALDRCYAAYRKTIQNOWK 282
QY 204 KEGNDKFTLNDFGFEMIHFSPYCKLVOKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVK 263
Db 283 QAGNNQPFLLDDQYMIFHTPFCKMVOKSLARLMFNDFLSASSDKQNNLYKGLEAFKGLK 342
QY 264 LEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNONGNMYTSSVYGSILASVLAQYSPQ 323
Db 343 LEETYNKQVDKALKLAKASLDMFNKTKASLYLSTNGNMYTSSLYGCLASLSHSAQEL 402
QY 324 AGKRIGVFSYSGSLAATLYSLKVTODATPGSALDKITASLCDLKSRLDSRTGVAPDVF 383
Db 403 AGSRIGAFSYSGSLAASFSSFRVSKDASPGSPLEKLVSSVSDLPKRLDSRRRMSPEEFTE 462
QY 384 NMLKREDTHLNVNIPQGSIDSLFEGTWYLVRVDEKGRRTYARRP 428
Db 463 IMNQREQFYHKVNFSPPGDTNSLFPGTWYLERVDEQHRKRYARRP 507

RESULT 8

B55729 hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5), mitochondrial - mouse
C/Species: Mus musculus (house mouse)
C/Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 05-May-2000
C/Accession: B55729
R/Boukafane, Y.; Duncan, A.; Wang, S.; Labuda, D.; Robert, M.F.; Sarrazin, J.; Schappert
Genomics 23, 552-559, 1994
A/Title: Human mitochondrial HMG CoA synthase: liver cDNA and partial genomic cloning, c
A/Reference number: A55729, MUID:95154824, PMID:7851882
A/Accession: B55729
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-471 <BOU>
A/Cross-references: GB:U12790, GB:U12791
C/Superfamily: hydroxymethylglutaryl-CoA synthase
C/Keywords: carbon-carbon lyase; coenzyme A; mitochondrion; oxo-acid-lyase

Query Match 59.5%; Score 1493; DB 2; Length 471;
Best local similarity 59.4%; Pred. No. 1.6e-103;
Matches 276; Conservative 74; Mismatches 73; Indels 42; Gaps 1;

QY 6 PLNAEACWPKDVGIVALEIYFPPSQYVDQAELEKYDGVDAKYYTIGLGOAKMGFCTDREDI 65
Db 6 PLAKTDTWPKDVGILALEVYFPAQYVDQDTLEKFNVNEAGKYTVGLGQTRMGFCSVQEDI 65
QY 66 NSLCMTVVQNLMERNNLSDYDCIGRLLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEG1DT 125
Db 66 NSLCITVQRLMERITKLPMDAVGRLEVGTEITIIDKSKAVKTVLMELFQDSGNTDIEG1DT 125
QY 126 TNACYGGTAAVFNAVNMIESSWD----- 149
Db 126 TNACYGGTASLFNNAANMMESSYWDGRYALVVCGDIAVYPSGNARPTGAGAVAMLIGPKA 185
QY 150 -----GLRGTHMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQ 203
Db 186 PLVLEQGLRGTHMENAYDFYKPNLASEYPIVDGKLSIQCYLRALDRCYAAYRKTIQNOWK 245
QY 204 KEGNDKFTLNDFGFEMIHFSPYCKLVOKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVK 263
Db 246 QAGNNQPFLLDDQYMIFHTPFCKMVOKSLARLMFNDFLSASSDKQNNLYKGLEAFKGLK 305
QY 264 LEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNONGNMYTSSVYGSILASVLAQYSPQ 323

Db 306 LEETYNKDVDKALLKASLDWMFNQKTKASLYLSTNNGNMVYSSLYGCLASLSHSAQEL 365
QY 324 AGKRIGVFSYSGSLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAE 383
Db 366 AGSRIGAFSYSGSLAASFSPFRVSKDASPGSPLEKLVSSVSDLPKRLDSRRRMSPEEFTE 425
QY 384 NMKLREDDTHLVNYPQGSIDSLFEGTWYLVRVDEKRRRTYARRP 428
Db 426 IMNQREQFYHKVNFSPPGDTSNLPFGTWYLERVDEMHRKRYARCP 470

RESULT 9
S38986
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - German cockroach
C:Species: Blattella germanica (German cockroach)
C:Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C:Accession: S38986
R;Martinez-Gonzalez, J.; Buesa, C.; Piulachs, M.D.; Belles, X.; Hegardt, F.G.
Eur. J. Biochem. 217, 691-699, 1993
A:Title: 3-Hydroxy-3-methylglutaryl-coenzyme-A synthase from Blattella germanica. Clonin
A:Reference number: S38986; MUID:94039108; PMID:7901012
A:Accession: S38986
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-453 <MAR>
A:Cross-references: EMBL:X73679; NID:g416168; PIDN:CAA52032.1; PID:g1772495
C:Superfamily: hydroxymethylglutaryl-CoA synthase
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 53.9%; Score 1352.5; DB 2; Length 453;
Best Local Similarity 56.6%; Pred. No. 4.7e-93;
Matches 259; Conservative 66; Mismatches 84; Indels 49; Gaps 4;
QY 13 WPKDVGIVALEIYFP SQYVDQAELEKYDGVDAKYYTIGLQAKMGFCTDREDINSLCMTV 72
Db 2 WPSDVGIVALEIYFP SQYVDQAELEKYDGVDAKYYTIGLQAKMGFCTDREDINSLCMTV 61
QY 73 VQNMERNNLSYDCIGRLEVGTEETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYG 132
Db 62 VSRIMERWSIPYSQIGRLEVGTEETIIDKSKSVKTNLMQLFKD--NTDIEGVDTVNACYG 119
QY 133 TAAVFNAVNWIESSWD-----G 150
Db 120 TSALEFNAISWVESSSWDGRYALVVAADIAVYAKGSARPTGAGAVAMLVGANAPLVFDRG 179
QY 151 LRGTMOHAYDYFKPDMLSYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDK 210
Db 180 VRSSHMQHAYDYFKPDLSSLYPYDGKLSIQCYLSALDHCTQLYCSKI---QKQLGEK- 234
QY 211 FTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEAFGVDKLEDTYFD 270
Db 235 FDIERLDVLFHAPYCKLVQKSLARMLNDFVRASEERTTKYSSLEALKGVKLEDITYFD 294
QY 271 RDVEKAFMKASSELFSQKTKASLIVSNQNGNMYTSSVYGSILASVLAQYSPQOLAGKRIGV 330
Db 295 REVEKAVMTYSKNMFEKTKPSLLANQVGNMYTTPSLYGLVSLVSKSAQELAGKRYAL 354
QY 331 FSYSGSLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLRED 390
Db 355 FSYSGSLAASMFSLRISSDASAKSLQRLVSNLSHIKPOLDLRHKVSPEEFAQTMETREH 414
QY 391 TTHLVNYPQGSIDSLFEGTWYLVRVDEKRRRTYARRP 428
Db 415 NNHKAQPTPEGSIDVLFPGTWYLESVDSLRYRSYQVP 452

RESULT 10
A53565
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - German cockroach
C:Species: Blattella germanica (German cockroach)
C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 21-Jul-2000
C:Accession: A53565

R;Buesa, C.; Martinez-Gonzalez, J.; Casals, N.; Haro, D.; Piulachs, M.D.; Belles, X.; Heg
J. Biol. Chem. 269, 11707-11713, 1994
A:Title: Blattella germanica has two HMG-CoA synthase genes. Both are regulated in the o
A:Reference number: A53565; MUID:94216267; PMID:7909314
A:Accession: A53565
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-455 <BUE>
A:Cross-references: GB:X77516; NID:g488133; PIDN:CAA54652.1; PID:g488134
C:Superfamily: hydroxymethylglutaryl-CoA synthase
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 50.7%; Score 1272.5; DB 2; Length 455;
Best Local Similarity 53.7%; Pred. No. 4.4e-87;
Matches 246; Conservative 64; Mismatches 101; Indels 47; Gaps 4;
QY 11 ACWPKDVGIVALEIYFP SQYVDQAELEKYDGVDAKYYTIGLQAKMGFCTDREDINSLCM 70
Db 2 AHWPEDVGIGIEMIFPSLYVDQAELETYDEVSPGKYTMGLGQDKMGVCTDREDINSLCL 61
QY 71 TVVQNMERNNLSYDCIGRLEVGTEETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACY 130
Db 62 TAVDKLMERNNIDYNDIGWLEVGTEETIIDKSKSVKTNLMQLFEESGNTDVEGIDTTNACY 121
QY 131 GGTAAVFNAVNWIESSWD----- 149
Db 122 RGTALFENALIWIESSSWDGRYALVVAADIAIYAKECSPGTGAGALLMLIGANAPIVDR 181
QY 150 GLRGTHMOHAYDYFKPDMLSYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDK 209
Db 182 GYVASHMKHAYDYFKPDMLSYPIVDGKLSVQCYLSALDHCTYPRFCSTEKYLKRCGKEN 241
QY 210 -DFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEAFGVDKLEDTY 268
Db 242 TKIDLDYFDFVFNHSPYCKLVQKSVARLVNDFI--QYPEK--YQDLQQLRNLKFEPTY 296
QY 269 FDRDVEKAFMKASSELFSQKTKASLIVSNQNGNMYTSSVYGSILASVLAQYSPQOLAGKRI 328
Db 297 FDRDIEKIFMDKSKQLFEKTKPSIMLANQVGNMYTTPSLYGLVSLISEDIGELAGKCI 356
QY 329 GVSYSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLR 388
Db 357 CMFSYSGSFAASMFSLHISTSSPGSTLSRLVTNLTHIKQVQQRVKLSPEGFENIMEIR 416
QY 389 EDTHVLVNYIPQGSIDSLFEGTWYLVRVDEKRRRTYAR 426
Db 417 EQNHKAQPTTPVAPSPNTLFPGTWYLESIDSMHRKRYKR 454

RESULT 11
T09688
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5), ozone-inducible - Scotch pine
N:Alternate names: 3-hydroxy-3-methylglutaryl-coenzyme A synthase
C:Species: Pinus sylvestris (Scotch pine)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T09688
R;Wegener, A.; Gimbel, W.; Werner, T.; Hani, J.; Ernst, D.; Sandermann, H.
Biochim. Biophys. Acta 1350, 247-252, 1997
A:Title: Molecular cloning of ozone-inducible protein from Pinus sylvestris L. with high
A:Reference number: Z16823; MUID:97214637; PMID:9061017
A:Accession: T09688
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-474 <MEG>
A:Cross-references: EMBL:X96386; NID:g1655678; PIDN:CAA65250.1; PID:g1655679
A:Experimental source: tissue-type needles
C:Function:
A:Description: catalyzes the condensation of acetyl coenzyme A (Ac-CoA) with acetoacetyl
tase
C:Superfamily: hydroxymethylglutaryl-CoA synthase
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase
Query Match 40.2%; Score 1009.5; DB 2; Length 474;

Best Local Similarity 43.5%; Pred. No. 1.9e-67;
Matches 208; Conservative 76; Mismatches 141; Indels 53; Gaps 7;

```

0Y 14 PKXGVIALEITYFPQYVDQAELEKYGVDAGKTYTIGLGOAKMGECTDREDINSLCMTVV 73
Db 5 PENVGILAMEIYFPTTCVQOEDLETFDVGSKGKTYTIGLGQDCMTCFTDLEDVISMSTLAV 64
0Y 74 QNLMERNNLSYDCIGRLEVGTEETIIDKSKSVKTNLMQLEESGNTDIEGIDTTNACYGCT 133
Db 65 TSULEKYEIDPKQIGRLEVGSETIVIDKSKSIKTWLMHIFKCGNTEIEGV DSTNACYGCT 124
0Y 134 AAVENAVNMWIESSWDG-----L 151
Db 125 AALFNCINWIESSSWDGRYGLVVAATDSAVYAEGAARPTGGAAVAMLJGNAPATESKY 184
0Y 152 RGTMOHAYDYFYKPDMLSEYPYVDGKLSIQCYL.SALDRCYSVYCKKIHQWQKEGNDKDF 211
Db 185 RGTMAHAYDYFYKPNLASEYPAVVDGKLSQTCYLMALDSCYKRFCKNKE--KEEG--RQF 239
0Y 212 TLNDFGFMIFHSPYCKLVQKSLARMNLNDFLNDQNRDKNSIYSGLEAFGDKLEDYFDR 271
Db 240 SLDDTYIAFHSPYKNKLVOQKSFGRLLFNDFSRHARSYGKDAQEKLPEPAGLSEQDSYNSR 299
0Y 272 DVEKAFMKASSELFSQKTKASLIVSNQNGNMYTSSVYGSIALVLAQYSPOQLAGKRIGVF 331
Db 300 DLEKVSQQLAKPLDYDAKIQPSTLLPKQVGNMITYASLYAALASII-HNKHTTLDGQRVMMF 358
0Y 332 SYGSGLAATLYSLKVTO DATPGSALDKITASL.CDLKSR.LDSRTGVADVFAENMKLREDT 391
Db 359 SYGSGLASTLFSFKIREGQFP-FTLSNIT-EVMDVQNKLD.SRHEFLPEDFVENTLKRMETL 416
0Y 392 HHLVNYIPQGSIDSLFEGTWTYLVVRVDEKHRRTYARPTPNDDTLDEGV--GLVHSNT 446
Db 417 YGAKDFVSTQSLRLRPGAFYLT.KVDSMYRRFRYSRKVISAGDNFEKSKLANGTTHDEL 474

```

RESULT 12
 T49718
 probable hydroxymethylglutaryl-CoA synthase [imported] - *Neurospora crassa*
 N:Alternate names: protein B23L21.310
 C:Species: *Neurospora crassa*
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jun-2000
 C:Accession: T49718
 R:Schulte, U.; Aign, V.; Heheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, May 2000
 A:Reference number: Z25022
 A:Accession: T49718
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-454 <SCH>
 A:Cross-references: EMBL:AL356172; GSPDB:GN00116; NCSP:B23L21.310
 A:Experimental source: BAC clone B23L21; strain OR74A
 C:Genetics:
 A:Gene: NCSP:B23L21.310
 A:Map position: 6
 A:Introns: 20/3; 55/1; 409/3
 C:Superfamily: hydroxymethylglutaryl-CoA synthase

	Query Match	Best Local Similarity	Score	DB 2;	Length
	Matches	Conservative	73;	Mismatches	Indels
	207;		983.5;	115;	69;
			44.6%;		Gaps
			Pred. No. 1,6e-65;		9,
QY	14	PKDVGIVALEITYFPSQYVDAQBLEEKYDGVADGAKYTIIGLGQAKMGFCTDRREDINSLCMTVV	73		
Dd	5	PONIGIKALIEIFYEPSQYVEQESELKEKFDGVS TGKYTII GLGTAKAFCCDDREDIYSLALTAV	64		
QY	74	QNLMERNNISYDCIGRLEVGTEETIIDKSRSVKTNLMQLFEESGNTDIEGIDITNACYGGT	133		
Dd	65	SRLLNKEYIEDTNTIGRLEVGTEETLLDKSKSVKSVLMOLFGE--NTNIEGVDITNACYGGT	122		
QY	134	AAVFENAVNTIESSWD-----GL	151		
Dd	123	NAFENSVWMIESSAWMDGRDAIVAGDIALYAKGNARPTGCAGCVAMLVGPENAPIAVERPGL	182		

[illegible]

RESULT 13
T09341
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Arabidopsis thaliana
N/Alternate names: 3-hydroxy-3-methylglutaryl-coenzyme A synthase; protein T26M18.30
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 20-Jun-2000
C/Accession: T09341; JC4567
R/Beyan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, June 1999
A/Reference number: Z16650
A/Accession: T09341
A/Molecule type: DNA
A/Residues: 1-461 <BEV>
A/Cross-references: EMBL:AL078606
A/Experimental source: cultivar Columbia; BAC clone T26M18
R/Montamat, F.; Guilloton, M.; Karst, F.; Delrot, S.
Gene 167, 197-201, 1995
A/Title: Isolation and characterization of a cDNA encoding Arabidopsis thaliana 3-hydroxy
A/Reference number: JC4567; MUID:96144274; PMID:8566777
A/Accession: JC4567
A/Molecule type: mRNA
A/Residues: 1-305, 'S', 307-341, 'N', 343-461 <MON>
A/Cross-references: EMBL:X83882; NID:G1143389; PIDN:CAAS8763.1; PID:G1143390
C/Comment: This enzyme mediates the conversion of three acetyl-CoA molecules to one molec
vity, and uptakes amino acids.
C/Genetics:
A/Gene: ATSP:T26M18.30
A/Map position: 4
A/Intons: 21/3; 57/2; 101/3; 149/3; 202/3; 227/2; 252/3; 271/2; 300/3; 347/3; 397/3
C/Superfamily: hydroxymethylglutaryl-CoA synthase
C/Keywords: carbon-carbon lyase; coenzyme A; glycoprotein; oxo-acid-lyase
F.117/Active site: Cys (covalent substrate-binding) #status predicted
F.269/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	39.0%;	Score 980;	DB 2;	Length 461;
Best Local Similarity	43.3%;	Pred. No. 2.9e-65;		
Matches 198;	Conservative 79;	Mismatches 128;	Indels 52;	Gaps 6;
QY	15	KDVGIVALEIYFPSPQYVDQAEEKYDGDVDAKTYITGLGQAKMGCTDRDINSICTWVQ	74	
Db	3	KNVGIIAMDIFYPPTCVQQEALFAHDGASKGYTIGLGQDCIAFCTELEDVISMENAVT	62	
QY	75	NLMERNNLSYDCIGRLEVGTEETIIDKS SVKTNLMQLFEESGNTDIEGIDTTNACYGTA	134	
Db	63	SLEFKYKIDPNQIGRLEVGSSETVIDSKS IKTFLMQLFEKCGCNTDVEGVDS TNACYGTA	122	
QY	135	AVFNAAVMITESSWDG-----	-LR 152	
Db	123	ALLNCVMNVESNSWDGRYGLVICTDSA VYAEGPARPTGGAALIAMLIGPDAPIVE SKLR	182	
QY	153	GTHMQHAYDYFKPDMISEPIVDGKL SIQCYLSALDRCRYSVCKKIHAQMOKESNDK DFT	212	

Db 183 ASHMAHYVDYKPNLASEYPPVVDGKLSQTCYLMALDSCYKHLCKNKE--KIEG--KEFS 237
Qy 213 LNDFGFMI FHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRD 272
Db 238 INDADYIVFHSPLYNKLVOQSFARLNYDFLRNASSIDEAAKEKFTPYSSLTLDSEYQSRD 297
Qy 273 VEKAFMKASSELFSQKTASLIVSNQNGNMYTSSVYGSILASVLAQYSPQOLAGKRIGVFS 332
Db 298 LEKYSQOIAKPFYDAKVQPTTLIPKEVGNNYTAFLYAFAFLIHK-KHNDLAGKRVVMFS 356
Qy 333 YGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVA PDVFAENMKLREDTH 392
Db 357 YGSGSTATMFSRLRNDNKPFSISN--IASVMDVGKCLKARHEYAPEKFVETMKLMEHRY 414
Qy 393 HLVNYI--PQGSIDSLFEGTWYLVVRVDEKHRRTYARR 427
Db 415 GAKDFVTTKEGTIDLLAPGTYYLKEVDSLRYRFGKK 451

RESULT 14
S58202
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YM4987.09c; protein YML126c
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 05-May-2000
C:Accession: S58202
R:Bowman, S.
submitted to the EMBL Data Library, July 1995
A:Reference number: S58194
A:Accession: S58202
A:Molecule type: DNA
A:Residues: 1-491 <BOW>
A:Cross-references: EMBL:Z50178; NID:g927528; PIDN:CAA90557.1; PID:g927536; MIPS:YML126c
A:Experimental source: strain AB972
C:Genetics:
A:Gene: SGD:HMG5
A:Cross-references: SGD:S0004595; MIPS:YML126c
A:Map position: 13L
C:Superfamily: hydroxymethylglutaryl-CoA synthase
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 37.9%; Score 950.5; DB 2; Length 491;
Best Local Similarity 43.4%; Pred. No. 5.1e-63;
Matches 198; Conservative 72; Mismatches 127; Indels 59; Gaps 7;
Qy 14 PKDVGIVALEIYFPSQYVDQALEKXGDVDAGKYTIGLGQAKMGFCTDREDINSLCMTVV 73
Db 46 PQNVGIGKIQIYIPTQCVNQSLEKFDGVSQGKYTIGLGQTNMSFVNDREDIYSMSLTVL 105
Qy 74 QNLMERNNLSDYDCIGRLEVGETETIIDKSKSVKTNLMOLFEESGNTDIEGIDTNNACYGGT 133
Db 106 SKLIKSYNIDTNKIGRLEVGETETIIDKSKSVKSLMQLFGE--NTDVEGIDTLNACYGGT 163
Qy 134 AAVFNAVNWIESSSWDG-----LR 152
Db 164 NALFNLSLNMWIESNAMDGRDAIVCGDIAIYDKGAARPTGAGTVA MWIGPDAPIVFDSVR 223
Qy 153 GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQW--QKEGNDK 209
Db 224 ASYMEHAAYDFYKPDFTSEYPPYVDGHSILTCYKALDQVYKSYSKKAISKGLVSDPAGSDA 283
Qy 210 DFTLNDFGFMI FHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEA-FGDVKLEDTY 268
Db 284 LNVLKTFDYNVFHVPTCKLVTKSYGRLLYNDF---RANPOLFPEVDAELATRDYDESL 338
Qy 269 FDRDVEKAFMKASSELFSQKTASLIVSNQNGNMYTSSVYGSILASVLAQYSPQOLAGKRI 328
Db 339 TDKNIETKTFVNVAKPFHKERV AQSLIVPTNTGNNYTA SVYAAFA SLINLVGSDDLQGKRV 398
Qy 329 GVFSYSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVA PDVFAENMKLR 388
Db 399 GLFSYSGSLAASLYSCKIVGD-----VQHIIKEL-DITNKLAKRITETPKDYEAALIELR 451

Qy 389 EDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKHRRTY 424
Db 452 ENAHLLKNFKPQGSIEHLSQSGVYYLLTNIDDKFRRSY 487

RESULT 15
S61875
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 05-May-2000
C:Accession: S61875; T38843
R:Katayama, S.; Adachi, N.; Takao, K.; Nakagawa, T.; Matsuda, H.; Kawamukai, M.
Yeast 11, 1533-1537, 1995
A:Title: Molecular cloning and sequencing of the hcs gene, which encodes 3-hydroxy-3-methylglutaryl-CoA synthase
A:Reference number: S61875; MUID:96353436; PMID:8750242
A:Accession: S61875
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-447 <KAT>
A:Cross-references: EMBL:U32187; NID:g974430; PIDN:AAB17601.1; PID:g974431
A:Note: the authors translated the codon AGT for residue 236 as Arg, TCT for residue 358
R:Gentles, S.; Churcher, C.M.; Barrrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A:Reference number: Z21815
A:Accession: T38843
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-447 <GEN>
A:Cross-references: EMBL:Z98530; PIDN:CAB11060.1; GSPDB:GN00066; SPDB:SPAC4F8.14c
A:Experimental source: strain 972h-; cosmid c4F8
C:Genetics:
A:Gene: SPDB:SPAC4F8.14c
A:Map position: 1
A:Introns: 20/3
C:Superfamily: hydroxymethylglutaryl-CoA synthase
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 35.8%; Score 900; DB 2; Length 447;
Best Local Similarity 41.5%; Pred. No. 2.6e-59;
Matches 191; Conservative 72; Mismatches 133; Indels 64; Gaps 6;
Qy 15 KDVGIVALEIYFPSQYVDQALEKXGDVDAGKYTIGLGQAKMGFCTDREDINSLCMTVVQ 74
Db 6 KDIGIKGLVLYTPNQYVEQALEAHADGVSTGKYTIGLGITKA FVDDRREDIYSFGLTALS 65
Qy 75 NLMERNNLSDYDCIGRLEVGETETIIDKSKSVKTNLMOLFEESGNTDIEGIDTNNACYGCTA 134
Db 66 QLIKRYQIDISKIGRLEVGETETIIDKSKSVKSLMQLFGD--NHNVEGIDCVNACYGGVN 123
Qy 135 AVFNAVNWIESSSWD-----GLR 152
Db 124 ALFNTIDWIESSAWDGRDGI VVAGDIALYAKGNARPTGAGCVALLVGPNAPIVFEPLR 183
Qy 153 GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKDF 212
Db 184 GTVMQHAYDFYKPDLTSEYPPYVDGHSILTCYKALDGA YANVNRDVA--KNGKSQGLG 240
Qy 213 LNDFGFMI FHSPYCKLVOKSLARMLNDFL---NDQNRDKNSIYSGLEAFGDVKLEDTY 268
Db 241 LDRFDYCI FHPCTKOVQKAYARLLYTDSAABEPSNPELBEVRELLSTLDA-----KKS 294
Qy 269 FDRDVEKAFMKASSELFSQKTASLIVSNQNGNMYTSSVYGSILASVLAQYSPQOLAGKRI 328
Db 295 TDKALEKGLMAITKERFNKRVS VYAPTNGNMYTASIFSCITALLSRVPADCLKGKRV 354
Qy 329 GVFSYSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVA PDVFAENMKLR 388
Db 355 GAYSYSGSLAASFFSFVVKGDVSE-----IAKTTNLVNDLDRHCLTPTQYEAEALIELR 407
Qy 389 EDTHHLVNYIPQGSIDSLFEGTWYLVVRVDEKHRRTYARRP 428
Db 408 HQAHLLKNFTPKSIERLRS GTYYLLTGIDDMFRRSYSYVKP 447

Fri Jun 25 07:33:56 2004

us-10-622-516-2.rpr

Page 8

Search completed: June 24, 2004, 13:21:29
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2004, 11:17:16 ; Search time 84 Seconds
(without alignments)
1795.449 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPGSPLPNAEACWPKDVGIV.....PRLPATAEPPEAAVISNGEH 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp Vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriaph:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	2403.5	95.7	509	4	Q8N995	Q8n995 homo sapien
2	2348	93.5	520	11	Q8JZK9	Q8jzk9 mus musculu
3	2344	93.3	520	11	Q8K0I5	Q8k0i5 mus musculu
4	2003	79.8	520	13	Q7ZYN7	Q7zyn7 xenopus lae
5	1923.5	76.6	508	13	Q7ZWE2	Q7zwe2 brachydanio
6	1494	59.5	508	4	Q8N7N8	Q8n7n8 homo sapien
7	1493	59.5	508	11	Q9DBM4	Q9dbm4 mus musculu
8	1491	59.4	508	11	Q9DBK1	Q9dbk1 mus musculu
9	1337	53.2	465	5	Q9V7N8	Q9v7n8 drosophila
10	1164	46.4	457	5	Q9NDA8	Q9nda8 dendroctonu
11	1047	41.7	463	10	Q8H051	Q8h051 oryza sativ
12	1044	41.6	450	3	Q9HE19	Q9he19 phycomyces
13	1029.5	41.0	268	11	Q8C5F4	Q8c5f4 mus musculu
14	1013	40.3	464	10	Q944F8	Q944f8 hevea bras
15	1009.5	40.2	474	10	P93773	P93773 pinus sylve
16	1009	40.2	464	10	Q94ET0	Q94et0 hevea bras

17	992	39.5	461	10	Q9M6U3	Q9m6u3 brassica ju
18	983.5	39.2	454	3	Q9P5J8	Q9p5j8 neurospora
19	978	38.9	468	5	Q86HL5	Q86hl5 dictyosteli
20	977.5	38.9	461	10	Q9FVG2	Q9fvq2 brassica ju
21	975.5	38.8	461	10	Q9FVG1	Q9fvq1 brassica ju
22	973.5	38.8	453	10	Q9FVG0	Q9fvq0 brassica ju
23	823	32.8	406	10	Q8L721	Q8l721 arabidopsis
24	445.5	17.7	382	5	Q8SR61	Q8sr61 encephalito
25	381	15.2	113	10	Q9M6A5	Q9m6a5 aerides jap
26	314	12.5	384	2	Q9FD66	Q9fd66 enterococcu
27	312.5	12.4	445	17	Q9HPJ0	Q9hpj0 halobacteri
28	306	12.2	389	2	Q9ZB67	Q9zb67 staphylococ
29	301.5	12.0	388	16	Q99R90	Q99r90 staphylococ
30	301	12.0	388	2	Q9FD82	Q9fd82 staphylococ
31	299.5	11.9	384	16	Q9CFA9	Q9cfa9 lactococcus
32	297.5	11.8	388	2	Q9FD87	Q9fd87 staphylococ
33	288.5	11.5	388	2	Q9FD76	Q9fd76 staphylococ
34	288.5	11.5	388	16	Q8CN06	Q8cn06 staphylococ
35	283	11.3	383	2	Q9FD71	Q9fd71 enterococcu
36	282	11.2	383	16	Q835L4	Q835l4 enterococcu
37	281.5	11.2	388	16	Q8EP69	Q8ep69 oceanobacil
38	272.5	10.9	388	16	Q92BU0	Q92bu0 listeria in
39	272	10.8	389	2	Q9KWG1	Q9kwg1 streptomyce
40	267.5	10.7	388	16	Q8Y781	Q8y781 listeria mo
41	254	10.1	388	2	Q8LI13	Q8li13 paracoccus
42	244	9.7	398	16	Q97PB3	Q97pb3 streptococ
43	243	9.7	398	2	Q9FD56	Q9fd56 streptococ
44	241	9.6	398	16	Q8DNS4	Q8dns4 streptococ
45	240	9.6	389	16	Q88VI4	Q88vi4 lactobacill

ALIGNMENTS

RESULT 1	ID	Q8N995	PRELIMINARY;	PRT;	509 AA.
AC	Q8N995;				
DT	01-OCT-2002 (TREMBlrel. 22, Created)				
DT	01-OCT-2002 (TREMBlrel. 22, Last sequence update)				
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)				
DE	Hypothetical protein FLJ38173.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RA	Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,				
RA	Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,				
RA	Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,				
RA	Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,				
RA	Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,				
RA	Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,				
RA	Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;				
RT	"NEDO human cDNA sequencing project."				
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AK095492; BAC04559.1; -.				
DR	GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.				
DR	GO; GO:0006084; P:acetyl-CoA metabolism; IEA.				
DR	InterPro; IPR008260; HMG-CoA synth.				
DR	Pfam; PF01154; HMG-CoA_synt; 1.				
KW	Hypothetical protein.				
SQ	SEQUENCE 509 AA; 56238 MW; 75AAB049EAB1460 CRC64;				

Query Match	95.7%;	Score 2403.5;	DB 4;	Length 509;
Best Local Similarity	89.8%;	Pred. No. 2.2e-168;		
Matches 467;	Conservative 0;	Mismatches 0;	Indels 53;	Gaps 2;
Oy	1	MPGSPLPNAEACWPKDVGIVALEIYFPSQYVDQAELEKYGVDAGKYTIGLGQAKMGFCT	60	
Db	1	MPGSPLPNAEACWPKDVGIVALEIYFPSQYVDQAELEKYGVDAGKYTIGLGQAKMGFCT	60	

QY 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
DB 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY 121 EGDITTNACYGCTAAVFNAVNWISSSSWD----- 149
DB 121 EGDITTNACYGCTAAVFNAVNWISSSSWD----- 149
QY 150 -----GLRGTHMQHAYDFYKPPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
DB 170 IGENAPLIFERGLRGTHMQHAYDFYKPPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 229
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
DB 230 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 289
QY 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSILASVLAQY 318
DB 290 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSILASVLAQY 349
QY 319 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
DB 350 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 409
QY 379 DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDTLLDEG 438
DB 410 DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDTLLDEG 469
QY 439 VGLVHSNIATEHIIPSPAKKVPRLPATAEPEAAVISNGEH 478
DB 470 VGLVHSNIATEHIIPSPAKKVPRLPATAEPEAAVISNGEH 509
RESULT 2
Q8JZK9 PRELIMINARY; PRT; 520 AA.
AC Q8JZK9;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to 3-hydroxy-3-methylglutaryl-coenzyme A synthase 1
DE (Hypothetical protein) (Pre B-cell leukemia transcription factor
DE 1).
GN B130032C06RIK OR HMGCS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Retina;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; BC029693; AAH29693.1; -
DR EMBL; BC034317; AAH34317.1; -
DR EMBL; BC023851; AAH23851.1; -
DR EMBL; AK031297; BAC27338.1; -
DR EMBL; AK044835; BAC32112.1; -
DR EMBL; AK045094; BAC32218.1; -
DR MGD; MGI:107592; Hmgcs1.
DR GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR InterPro; IPR008260; HMG_CoA_synth.

DR InterPro; IPR000590; HMG_CoA_synth_AS.
DR Pfam; PF01154; HMG_CoA_synth_1.
DR PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW Hypothetical protein.
SQ SEQUENCE 520 AA; 57569 MW; E28F8772CF64D85C CRC64;
Query Match 93.5%; Score 2348; DB 11; Length 520;
Best Local Similarity 86.9%; Pred. No. 2,8e-164;
Matches 452; Conservative 13; Mismatches 13; Indels 42; Gaps 1;
QY 1 MGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELKYGVDAGKTTIGLGQAKMGFCT 60
DB 1 MGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELKYGVDAGKTTIGLGQAKMGFCT 60
QY 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
DB 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
QY 121 EGDITTNACYGCTAAVFNAVNWISSSSWD----- 149
DB 121 EGDITTNACYGCTAAVFNAVNWISSSSWD----- 149
QY 150 -----GLRGTHMQHAYDFYKPPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
DB 181 IGENAPLIFERGLRGTHMQHAYDFYKPPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240
QY 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
DB 241 RAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300
QY 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSILASVLAQY 318
DB 301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSILASVLAQY 360
QY 319 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
DB 361 SPQOLAGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
QY 379 DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDTLLDEG 438
DB 421 DVEAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKRRTYARRPTPNDHSLDEG 480
QY 439 VGLVHSNIATEHIIPSPAKKVPRLPATAEPEAAVISNGEH 478
DB 481 MGLVHSNIATEHIIPSPAKKVPRLPATAEPEAAVISNGEH 520
RESULT 3
Q8K0I5 PRELIMINARY; PRT; 520 AA.
AC Q8K0I5;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
DE HMGCS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031363; AAH31363.1; -
DR MGD; MGI:107592; Hmgcs1.
DR GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR InterPro; IPR008260; HMG_CoA_synth.
DR InterPro; IPR000590; HMG_CoA_synth_AS.
DR Pfam; PF01154; HMG_CoA_synth_1.
DR PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.


```
Db      2  |||||:|:|:|||||:|:|:|||||:|:|:|||||:|:|:|||||:|:|:|
QY      73  VQNLERNNLSDYDCIGRLVGTETIIDKSKSVKTNLMQLFEEGNTDIEGIDTTNACYG 132
Db      62  VQRLERNGLSYESVGRLEVGTETIIDKSKTKVLMQLFEEGNTDVEGVDTTNACYG 121
QY      133 TAAVFNAVNWIESSWD-----G 150
Db      122 TALLFNAVNWIESSWDGRYALVAGDIAYVATGARSARPTGAGAVAMLVGPNAPLAFERG 181
QY      151 LRGTMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVYCKIHAQWQEGNDKD 210
Db      182 LRGTMOHAYDFYKPDMSYEVVDGKLSIQCYLSALDQCYSYVYKNIHARWQREGTEGR 241
QY      211 FTLNDFGFMIFHSFYCKLVOKSLARMLNDFLNDQNRDKN- IYSGLEAFGDVKEIDTYF 269
Db      242 CSLEDFGFMVHSPYCKLVOKSLARMLNDFLCHPSRNMEGPFSGLEAFRDVKIEDTYF 301
QY      270 DRDVEKAFMKASSELFSQKTASLLVSNQNGNMYTSSVYGSLSAVLAQYSPQOLAGKRIG 329
Db      302 DRDVEKAFMKASSELFDCKTKASLLISNNGNMYTSPYGCILASVLAQHTPQOLAGKRIG 361
QY      330 VFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVA PDVFAENMKLRE 389
Db      362 VFSYSGSFAATLYSIKVTQDATPGSALDKLVSSLCDLPARLDSRQKVS PGVFAETMKLRE 421
QY      390 DTHLVNYPQGSIDSLFEGTWYLVVRVDEKRRTYARRPTPNDDTLDEGVGLVHSNIATE 449
Db      422 ETHLVNYPQGSVDLFPGTWYLVTRVDEKRRQYARRSMNDRPLE -AGLVSSMAAE 479
QY      450 HIPSPAKVPRLPATAPEPAVISNGEH 478
Db      480 HIPSPKKMPRIPTTAGPEVVVWSNGDH 508
```

RESULT 6

```
Q8N7N8  PRELIMINARY; PRT; 508 AA.
ID Q8N7N8
AC Q8N7N8;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ40785.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RA Oshima A., Takahashi-Fuji A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isegai T.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK098104; BAC05233.1; -.
DR GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR InterPro; IPR008260; HMG_COA_synth.
DR InterPro; IPR000590; HMG_COA_synth.
DR Pfam; PF01154; HMG_COA_synth_1.
DR PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW Hypothetical protein.
SQ SEQUENCE 508 AA; 56856 MW; 156DD32AF48084B4 CRC64;
```

Query Match 59.5%; Score 1494; DB 4; Length 508;
Best Local Similarity 59.4%; Pred. No. 1.9e-101;
Matches 276; Conservative 74; Mismatches 73; Indels 42; Gaps 1;

```
QY      6  PLNAEACPKDVGIALLEIYPSQYVDQALEKXYDVGADGKTYTIGLQAKMGFCTDREDI 65
Db      43  PLAKTDWPKDVGIALLEIYPPAQYVDQTDLEKFNNEAGKTYVGLQTRMGFCVQEDI 102
QY      66  NSLCMTVQNLERNNLSDYDCIGRLVGTETIIDKSKSVKTNLMQLFEEGNTDIEGIDT 125
Db      103  NSLCMTVQRLMERTKLPMDAVGRLEVGTETIIDKSKAVKTVLMELFQDSGNTDIEGIDT 162
QY      126 TNACYGTAAVFNAVNWIESSWD----- 149
Db      163 TNACYGTAASLFNAANWIESSYWDGRYALVVGCDIAYVPSGNARPTGAGAVAMLIGPKA 222
QY      150  -----GLRGTMOHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVYCKIHAQWQ 203
Db      223  PLVLEQGLRGTHMENAYDFYKPNLASEYPIVDGKLSIQCYLRALDRCYAAYRKTIQNWQ 282
QY      204  KEGNDKDFTLNDFGFMIFHSFYCKLVOKSLARMLNDFLNDQNRDKNISYSGLEAFGDVK 263
Db      283  QAGNNQPFLLDVQYMI FHTFPCKMVOKSLARMLNDFLSSSSDKQNNLYKGLEAFRGLK 342
QY      264  LEDYFDRDVEKAFMKASSELFSQKTASLLVSNQNGNMYTSSVYGSLSAVLAQYSPQOL 323
Db      343  LEETYNKDVKALIKASLDMENQKTASLYLSTNNGNMYTSSLYGCILASHSAQEL 402
QY      324  AGKRIGVFSYSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVA PDVFAE 383
Db      403  AGSRIGAFSGSGLAASFVSFRVSKDASGSTEKLVSVSVDLPKRLDSRRRMSPEEFTE 462
QY      384  NMKLREDTHLVNYPQGSIDSLFEGTWYLVVRVDEKRRTYARRP 428
Db      463  IMNQREQFYHKVNFSPGDTSNLFPGTWYLVVRVDEKRRRYARCP 507
```

RESULT 7

```
Q9DBM4  PRELIMINARY; PRT; 508 AA.
ID Q9DBM4
AC Q9DBM4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE 3-hydroxy-3-methylglutaryl-coenzyme A synthase 2.
GN HMGCS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Irawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
```


Db 2 SAMPEDVGI IAL E IYFPAQYVDQTELEEQYDGVSA GKYTVGLGQKMGFCNDREDINSLCL 61

Qy 71 TVQNLMERNNLSYDCIGRLVEGTETI IDKSKSVKTNLMQIFEESGNTDIEGIDTNNACY 130

Db 62 TVVDNLLTRYQVKPQEI GRLEVGTEI IDKSKSVKTNLMQIFEPHGVTDIEGIDTNNACY 121

Qy 131 GGTAAVENAVNWI ESSSWDG----- 150

Db 122 GGLPPSSTPSIAVESSSWNGRYALVVAADIAVYAKGPAPRTGATGAVAMLVGNAPLVVD 181

Qy 151 --LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVYCKKIHAQWQKEGND 208

Db 182 RRLKSSFFKGAYDFYKPDLTSEYPVVDGKLSIQCYLEALDKCYQVYCQKTEKLLQ----- 236

Qy 209 KDTFLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDKLEDTY 268

Db 237 KPVSI ESFDGILFHTPYCKLVQKSVGRLLALNDFVRE--GKPELHPDLEKFPANIQLKDSY 293

Qy 269 FDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQOLAGKRI 328

Db 294 FDRDVEKAFLSASNAFLQRTLPSSLIAANIGNMYTSSLYGGLVSVYSRSLLENLVDNRY 353

Qy 329 GVEFSYSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRDRTGAPVDFAEENMKR 388

Db 354 AMFSYSGGLASSFY SITIKNNA-----ALQTLKNMLSVEPLAKRSKIAPKEFEATLELR 409

Qy 389 EDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKGRRTYAR 426

Db 410 QQSAHKVPEYEPVGDISHFPFGTFVLTKIDEQHRRVYDR 447

RESULT 11

Q8H051 ID Q8H051 PRELIMINARY; PRT; 463 AA.

AC Q8H051; 01-MAR-2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Putative hydroxymethylglutaryl coenzyme A synthase.

GN OJ1263H11.4.

OS Oryza sativa (japonica cultivar-group).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzae; Oryza.

OC NCBI_Taxid=39947;

OX [1]

RN RP SEQUENCE FROM N.A.

RA Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Currie J.,

RA Collura K., McCombie D.W.R., de la Bastide M., Spiegel L., Preston R.,

RA Kirchoff K., Kuit K., Nascimento L., Zutavern T., Balija V., Bell M.,

RA Baker J., Santos L., Miller B., Katzenberger F., Muller S., King L.,

RA Yang C., O'Shaughnessy A., Palmer L., Dedhia N.,

RT "Rice Genomic Sequence.";

RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC118980; AA015287.1; -.

DR GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.

DR GO; GO:0006084; P:acetyl-CoA metabolism; IEA.

DR InterPro; IPR008260; HMG_COA_synth.

DR InterPro; IPR000590; HMG_COA_synth_AS.

DR Pfam; PF01154; HMG_COA_synt; 1.

DR PROSITE; PS01226; HMG_COA_SYNTHASE; 1.

SQ SEQUENCE 463 AA; 51502 MW; 60D021C5C48E99B6 CRC64;

Query Match 41.7%; Score 1047; DB 10; Length 463;

Best Local Similarity 47.1%; Pred. No. 1.3e-68;

Matches 216; Conservative 70; Mismatches 111; Indels 62; Gaps 8;

Qy 15 KDVGIVALEIYFPSQYVDQAEL EKYDGDAGKYTIGLGQAKMGFCTDREDINSLCMTVVQ 74

Db 6 KDVGILAMDIYFPPTCVLQDELEHNDGVS KGYTIGLGQDSMAFCTEVEDVISMSLTIVK 65

Qy 75 NLMERNNLSDICIGRLVEGTETI IDKSKSVKTNLMQIFEESGNTDIEGIDTNNACYGTA 134

Db 66 SLEENYKIDPKCIGRLVEGSETVI DSKSIKTWLMQIFEEGNTDIEGVDSSNACYGTA 125

Qy 135 AVENAVNWI ESSSWDG-----LR 152

Db 126 ALFNCVNWVESNSWDGRYGLVVTDSAVYAEGPARPTGAAAIAMLIGPNAPVSFESKYR 185

Qy 153 GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVYCKKIHAQWQKEGNDKDF 212

Db 186 GSHMAHYVDFYKPDLA SEYPIVDGKLSIQCYLMALDSCVNYFCKKYE--KLEG--KQFS 240

Qy 213 LNDFGFMIFHSPYCKLVQKSLARMLLNDFL-----NDQNRDKNSIYSGLEAFGDKLED 266

Db 241 IHDADYFVHFSPYNKLVQKSFARLYYNDFLRKCTVEDGSRKLEPYSGLSS-----EE 294

Qy 267 TYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQOLAGK 326

Db 295 SYGSRLEKASQOVAKHLVDSKVQPIITLIPKQVGNMYTASLYAALASVMHNKN-ETLAGQ 353

Qy 327 RIGVFSYSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRDRTGAPVDFAEENMK 386

Db 354 RIVFSYSGGLTSTMFSEFKINEGQHP-FILSNI-AGILDVSKLLESRHVAPEKFVAALK 411

Qy 387 LREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKGRRTYA 425

Db 412 LMEHRYGAKDFTTSQDTSILAPGTYYLLTHVDSMYRRFYA 450

RESULT 12

Q9HE19 ID Q9HE19 PRELIMINARY; PRT; 450 AA.

AC Q9HE19; 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Putative 3-hydroxy-3-methylglutaryl coenzyme A synthase.

GN HMGS.

OS Phycomyces blakesleeanus.

OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;

OC Phycomyces.

OC NCBI_Taxid=4837;

OX [1]

RN RP SEQUENCE FROM N.A.

RC STRAIN=NRRL1555;

RA Ruiz-Albert J., Cerdà-Olmedo E., Corrochano L.M.;

RT "Genes for the metabolism of 3-hydroxy-3-methylglutaryl coenzyme A in the fungus Phycomyces.";

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ297414; CAC18553.1; -.

DR GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.

DR GO; GO:0006084; P:acetyl-CoA metabolism; IEA.

DR InterPro; IPR008260; HMG_COA_synth.

DR Pfam; PF01154; HMG_COA_synt; 1.

SQ SEQUENCE 450 AA; 50136 MW; 92B6E495039F0967 CRC64;

Query Match 41.6%; Score 1044; DB 3; Length 450;

Best Local Similarity 47.5%; Pred. No. 2.1e-68;

Matches 217; Conservative 68; Mismatches 116; Indels 56; Gaps 6;

Qy 13 WPKDVIVALEIYFPSQYVDQAEL EKYDGDAGKYTIGLGQAKMGFCTDREDINSLCMTV 72

Db 7 YPENVGILALEMYFPSRCVEQTA MEVYDGVSTGKYTIGLGQDKMAFIDREDIQSICLTA 66

Qy 73 VQNLMERNNLSYDCIGRLVEGTETI IDKSKSVKTNLMQIFEESGNTDIEGIDTNNACYG 132

Db 67 VHNLMKEYNIAITYDIGRLVEGTETI IDKSKSVKTTIMTLFAEHGNYEIEGIDTNNACYGG 126

Qy 133 TAAVENAVNWI ESSSWD-----G 150

Db 127 FSAFSAVAVNWI ESSSWDGRYAIVIAGDLALYASGAARPTSGAGVAVMLVGKDAPIVLGRG 186

Qy 151 LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSYVYCKKIHAQWQKEGNDKD 210

Db 187 LRYTYMDHAYDFYKPDMSSEYPIVDGKFSNVCIYRAFADACYNMYMKRLSKTIKQDVT--- 243

QY	211	FTLNDGFMIFHSPYCKLVOKSLARMLNDFLNDQNDKNSIYSGLEAFGDVKLEDIYED	270
Db	244	-SWEDVDVVCCHSPYAKLVTKSFARSAYNDFMADK---KNPKYALAPFEELAYEASLEN	299
QY	271	RDVEKAFMKASSELEFSQKTASLLVSNQNGNMNTSSYYGSLASVLAQYSPQOLAGKRIGV	330
Db	300	RDEKATATLTAKGYAQKVGAAYAPKQIGNMYAGAVWAGLASLVEVEVDTLKDKRYLL	359
QY	331	F5YSGSLAATLYSLKYTDATPGSALDKITASI.CDLKSRLLDSRTGVAADVFAENMKLRED	390
Db	360	YSYSGSLAASWVSFRVV-----GSTAD-1KAKL-NLRQLDARTHSKEAFAEAMQIREN	412
QY	391	THHLVNYIPOGSDSLFEGTWYLVAVDEKHKRTYARR	427
Db	413	THNACNYPFGSLIEHIAFGAYYVDKIDDKWRRFFYRK	449

ID	Q8C5F4	PRELIMINARY;	PRT;	268 AA.
DT	01-MAR-2003	(TREMBLrel. 23, Created)		
DT	01-MAR-2003	(TREMBLrel. 23, last sequence update)		
DE	01-OCT-2003	(TREMBLrel. 25, last annotation update)		
GN	HMGC51 OR B130032C06RIK.	Pre B-cell leukemia transcription factor 1.		
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RX	MEDLINE=22354683; PubMed=12466851;			
RA	The FANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RT	60,770 full-length cDNAs.";			
RL	Nature 420:563-573 (2002).			
DR	EMBL; AK078743; BAC37373.1; -.			
DR	MGD; MGI:107592; Hmgs1.			
DR	GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.			
DR	GO; GO:0006084; P:acetyl-CoA metabolism; IEA.			
DR	InterPro; IPR008260; HMG_CoA_synth.			
DR	InterPro; IPR000590; HMG_CoA_synth_AS.			
DR	Pfam; PF01154; HMG_CoA_synth; 1.			
DR	PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.			
CO	SEQUENCE 268 AA; 29493 MW; 9F6A7BBAFD4317F4 CRC64;			

Query Match	41.0%;	Score 1029.5;	DB 11;	Length 268;
Best Local Similarity	72.7%;	Pred. No. 1.1e-67;		
Matches 200; Conservative	10;	Mismatches 14;	Indels 51;	Gaps 2;

QY	I	MPSLPLNAEACMPKDVIGIVALEIYPSPQYVDQAELBEKYDGDVADGAKYITGLGOAKMGFCT	60
Dd	1	MPSLPLNAEACMPKDVIGIVALEIYPSPQYVDQAELBEKYDGDVADGAKYITGLGOARMGFCT	60
QY	61	DREDINSLCMTFYQNLMERNNLSDYCIGRLLEVGTETIIDKSXSVKTINMQLFEESGNTDI	120
Dd	61	DREDINSLCLTFVQKLMEHSHLSYDCIGRLEVGTEITIIDKSXSVKSNIMQLFEESGNTDI	120
QY	121	EGIDTTNACYGGTAAVENAVNMIBSSWD-----	149
Dd	121	EGIDTTNACYGGTAAVENAVNMWESSWDGRYALVVA GDIALYATGNARPTGVGAVALL	180
QY	150	-----GLRGTHMOHAYDEPKPDMISEYPIVDGLSIQCYSALDRCYSVYCKKI	198
Dd	181	IGPNAPLIFDRGLRGTHMOHAYDEPKPDMISEYPIVDGKLSIQCYISALDRCYSVYRKKI	240
QY	199	HAOWOKEGNDKDFTLNDFGFMIFHSPYCKLVOKSL	233
Dd	241	RAOWOKGG-----FSFSQBSCTRCWCMLGCSV	266

RESULT 14	
Q944F8	
ID Q944F8	PRELIMINARY; PRT; 464 AA.
AC Q944F8;	
DT 01-DEC-2001 (TREMBLrel. 19, Created)	
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)	
DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)	
DE Hydroxymethylglutaryl coenzyme A synthase.	
OS Hevea brasiliensis (Para rubber tree).	
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;	
OC eurosids I; Malpighiales; Euphorbiaceae; Crotonoideae; Micrandreae;	
OC Hevea.	
OX NCBI_TaxID=3981;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=cv, RIM600;	
RA Hallahan D.L., Keiper-Hrynko N.M.;	
RT "Genes involved in the biosynthesis of isopentenyl diphosphate in the	
RT rubber tree <i>Hevea brasiliensis</i> .";	
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.	
DR EMBL; AF429389; AAL18930.1; -	
DR GO; GO:0004421; F-hydroxymethylglutaryl-CoA synthase activity; IEA.	
DR GO; GO:0006084; P-acetyl-CoA metabolism; IEA.	
DR InterPro; IPR008260; HMG CoA synth.	
DR InterPro; IPR000590; HMG_CoA_synth_AS.	
DR Pfam; PF01154; HMG CoA synth. 1.	
DR PROSITE; PS01226; HMG CoA SYNTHASE; 1.	
SQ SEQUENCE 464 AA; 51157 MW; 980060294CBB5187 CRC64;	

Query Match	40.3%;	Score 1013;	DB 10;	Length 464;
Best Local Similarity	44.2%;	Pred. No. 4.2e-66;		
Matches 207; Conservative	86;	Mismatches 121;	Indels 54;	Gaps 8;

```

QY 15 KDVGIVALEIYFBSQYVDQAELEKDYGDAGKYTTIGLGQAKNGFCTDREDINSLCMTVVQ 74
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3 KNVGILAVDIFYFPPIYVQOEALFAHDGASKKYTTIGLGQDCMAFCTEVEDVISMLTAVT 62
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 75 NLMERNNLSTYDCIGRLVGTETIIDKSKSVKTNLMOLFEESENTDIEGIDTTNACYGGA 134
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 63 SLDDKYNIDPKQIGRLVGESETVIDSKSIKTFLMQIFKEFGNTDIEGVDSTNACYGGA 122
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 135 AVFNVAWNWIESSWDG-----LR 152
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 123 ALFNCVNWVESSWDGKRYGLVVCDSAYVAEGPARPTGAAAIAILVGPDAPIAFESKFR 182
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 153 GTHMQAHAYDFYKPPMLSEYPIVDKLSIQCYLSALDRCSYVCKKIHAQWQK--EGNDKDF 211
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 183 GSHMSHAYDFYKPNLASEYPVVDGKLSQTCYIMALDSCYKHF-----AKYEKFEK--KQF 236
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 212 TLENDFGFMIFHSPYCKLVOKSLARMLNDFLNDQNRDKNSIYSGLEAFGVDKLEDDTYFDR 271
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 237 SISDAEYFVFHSPYKNLVOKSFARLVFNDFVRNASSIDETAKEKLAFPSNLGDSBYQNR 296
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 272 DVEKAFMKASSELFSQKTKASLIVSNQNGNMYYTSSVYGSLASVLQAQYSPQOLAGKRIGVF 331
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 297 DLEKVSQOVAKPLVDAKVKPTTLIPKQVGNMYTASLYAFAFSL--HSKHTELAGKRVTLF 355
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 332 SYGSGLAATLYSLKVTDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLREDT 391
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 356 SYGSGLTATMFSLRLHEGQHPFS--LSNI--ASVMNVAGKUKARHELPRPEKFVDIMKLMHR 413
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 392 HHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTENDTLDGCV 439
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 414 YGAKDFTVTSKDCSLLASGTYLLTEVDLSYRRRPFYAQKAVGN--TVENG 459
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 15	
P93773	
ID P93773	PRELIMINARY;
AC P93773;	PRT; 474 AA.

DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE 3-hydroxy-3-methylglutaryl-CoA-synthase.
OS Pinus sylvestris (Scots pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3349;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Needle;
RX MEDLINE=97214637; PubMed=9061017;
RA Wegener A., Gimbel W., Werner T., Hani J., Ernst D., Sandermann H.;
RT "Molecular cloning of ozone-inducible protein from Pinus sylvestris L.
RT with high sequence similarity to vertebrate 3-hydroxy-3-methylglutaryl-
RT CoA-synthase.";
RL Biochim. Biophys. Acta 1350:247-252 (1997).
DR EMBL; X96386; CAA65250.1; -.
DR PIR; T09688; T09688.
DR GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR InterPro; IPR000886; ER_target_S.
DR InterPro; IPR008260; HMG_CoA_synth.
DR Pfam; PF01154; HMG_CoA_synth; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 474 AA; 52998 MW; 3ECB48E5F23D37BA CRC64;

Query Match 40.2%; Score 1009.5; DB 10; Length 474;
Best Local Similarity 43.5%; Pred. No. 7.8e-66;
Matches 208; Conservative 76; Mismatches 141; Indels 53; Gaps 7;

QY 14 PKDVGI~~VALEIY~~FP~~SYVDQAELEKYD~~GV~~DAGK~~YTTIG~~QAKMG~~CT~~DR~~EDINSL~~CM~~TVV 73
Db 5 PENVGILAMEIYFPTTCVQ~~QEDLETF~~DGVS~~KGK~~YTTIG~~QDCMT~~FCTDL~~EDV~~ISMSLTAV 64
QY 74 QNLMERNNLSYDCIGRL~~EVGTETI~~IDS~~KS~~SVKTNLMQLEESGNTDIEGIDT~~NA~~CYGGT 133
Db 65 TSLLEKYEIDPKQIGRL~~EV~~GESE~~TV~~IDS~~KS~~IKTWLMHFEKCGNTEIEGVDST~~NA~~CYGGT 124
QY 134 AAVFN~~AVNW~~IESSSWD~~G~~-----L 151
Db 125 AALFNCINWIESSSWDGRYGLV~~AT~~DSAVYAEGAARPTGAA~~AV~~AMLIGPNAPIATESKY 184
QY 152 RGT~~MO~~HAYD~~FYK~~PDMLSEYPIVDGKLSIQCYLSALDR~~CYS~~VYCKIHAQWQKEGNDKDF 211
Db 185 RGT~~MA~~HVYD~~FYK~~PNLASEYVVDGKLSQTCYLMALDSCYK~~RF~~CNKFE--KEEG--RQF 239
QY 212 TLNDFGMI~~FH~~SPYCKLVOKSLARMLND~~FL~~NDQNRDKNSIYSGLEAFG~~DV~~KLEDTYFDR 271
Db 240 SL~~LD~~TDYIA~~FH~~SPY~~NK~~LVQKSGFRLL~~FN~~DFS~~R~~HARSVGKDAQE~~KLE~~PFAGLSEQDSYNSR 299
QY 272 DVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGS~~LA~~SVLAQYSPQQLAGKRIGVF 331
Db 300 DLEKVSQQLAKPL~~Y~~DAKIQPSTLLPKQVGNMYTASLYAALASII-HNKHTTLDGQRYMMF 358
QY 332 SYGSGLAATLYSLKVTQDA~~TP~~GSALDKITASLCDL~~KSR~~LD~~SRT~~GVAPDVF~~AE~~NMKLR~~ED~~T 391
Db 359 SYGSGLASTLFSFKIREGQFP-FTLSNIT-EVMDVQNKLD~~SR~~HEFLPEDFVENL~~KR~~METL 416
QY 392 HHLVNYIPQGSIDSLFEGTWYLV~~R~~VEDEK~~HR~~RTYARRPT~~ND~~TLDEGV--GLVHSNI 446
Db 417 YGAKDFVSTSQLSLRPGAFYLT~~KV~~DSMYRRFYSRKVISAGDNFEKSKLANGTTHDEL 474

Search completed: June 24, 2004, 13:19:30
Job time : 88 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 24, 2004, 13:28:14 ; Search time 2861 Seconds
(without alignments)
4989.207 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPEGLPLNAEACWPKDVGV.....PRLPATAEPEAAVISNGEH 478

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10622516/runat_23062004_162633_660/app_query.fasta_1.647
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10622516 @CGN_1_1_5180 @runat_23062004_162633_660 -NCPU=3 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2348	93.5	3222	11	AK045094	AK045094 Mus muscu
2	2348	93.5	3286	11	AK044835	AK044835 Mus muscu
3	2348	93.5	3466	11	AK031297	AK031297 Mus muscu
4	1513	60.3	1527	29	AY416670	AY416670 Homo sapi
5	1493	59.5	1527	29	AY416672	AY416672 Mus muscu
6	1493	59.5	3287	11	AK004865	AK004865 Mus muscu
7	1491	59.4	2524	11	AK004902	AK004902 Mus muscu
8	1382	55.0	1083	13	BX441322	BX441322 BX441322
9	1285	51.2	1420	29	AY416671	AY416671 Pan trogl
10	1263	50.3	998	12	BM474567	BM474567 AGENCOURT
11	1263	50.3	1201	13	BX419944	BX419944 BX419944
12	1255.5	50.0	905	13	BU508388	BU508388 AGENCOURT
13	1247	49.7	789	14	CA512392	CA512392 UI-R-FUO-
14	1216	48.4	779	14	CD354310	CD354310 UI-M-GMO-
15	1210	48.2	967	13	BU149049	BU149049 AGENCOURT
16	1206	48.0	780	14	CA316122	CA316122 UI-M-FWO-
17	1197	47.7	960	12	BQ049460	BQ049460 AGENCOURT
18	1196	47.6	914	13	BU149022	BU149022 AGENCOURT
19	1173.5	46.7	964	13	BU514917	BU514917 AGENCOURT
20	1171.5	46.7	877	9	AU132233	AU132233 AU132233
21	1149	45.8	810	14	CD354238	CD354238 UI-M-GMO-
22	1147	45.7	937	13	BX441044	BX441044 BX441044
23	1146.5	45.7	1106	14	CD505314	CD505314 CDA73-A07
24	1143	45.5	1156	12	BM461079	BM461079 AGENCOURT
25	1138	45.3	1118	14	CD503546	CD503546 CDA63-A04
26	1136	45.2	914	13	BU941319	BU941319 AGENCOURT
27	1135	45.2	718	12	BM963690	BM963690 UI-M-BQO-
28	1134	45.2	1053	13	BU162936	BU162936 AGENCOURT
29	1133.5	45.1	1068	14	CD503357	CD503357 CDA61-H01
30	1132	45.1	881	13	BX853811	BX853811 BX853811
31	1117	44.5	681	14	CF169002	CF169002 B0807G11-
32	1115	44.4	866	13	BX742538	BX742538 BX742538
33	1112	44.3	725	14	CD354751	CD354751 UI-M-GMO-
34	1111	44.2	674	14	CA873594	CA873594 K0926C09-
35	1111	44.2	674	14	CF170711	CF170711 B0832B04-
36	1111	44.2	950	14	CF412994	CF412994 CH3#083_D
37	1108	44.1	682	14	CB288167	CB288167 CMD75_F05
38	1104	44.0	669	14	CA874751	CA874751 K0934F10-
39	1101	43.8	885	12	BG681747	BG681747 602627868
40	1099	43.8	823	12	BI871842	BI871842 603396981
41	1099	43.8	1201	13	BX445244	BX445244 BX445244
42	1093.5	43.5	756	14	CF737178	CF737178 UI-M-HD0-
43	1091	43.4	703	9	AU127534	AU127534 AU127534
44	1089	43.4	769	14	CD353582	CD353582 UI-M-GMO-
45	1087	43.3	657	14	CA874680	CA874680 K0934B05-

ALIGNMENTS

RESULT 1
AK045094
LOCUS
DEFINITION
AK045094 3222 bp mRNA linear HTC 20-SEP-2003
Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
enriched library, clone:BI30032C06 product:pre B-cell leukemia
transcription factor 1, full insert sequence.
ACCESSION
AK045094
VERSION
AK045094.1 GI:26337068
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, U., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001)
COMMENT Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gs.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
COMMENT Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hnako Atsugi City, Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge.
COMMENT Please visit our web site for further details.
URL: http://genome.gs.riken.go.jp/
URL: http://fantom.gs.riken.go.jp/.
FEATURES
source location/Qualifiers
1. .3222

CDS		/organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:B130032C06" /db_xref="MGI:2410453" /db_xref="taxon:10090" /clone="B130032C06" /tissue_type="parthenogenote" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="9.5 days embryo" 62.1624 /note="unnamed protein product; pre B-cell leukemia transcription factor 1 (MGD MGI:97495, GB NM_008783, evidence: BLASTN, 99%, match=124) putative" /codon_start=1 /protein_id="BAC32218.1" /db_xref="GI:26337069" /translation="MPGSLPLNMEACWPKDVGIVALEIYPSQYVDQAELEKYDGVDA GKVTIGLQARMGFCPTDREDINSLCLTVQKLMERHSLSYDCIGRIEFGTEITIDSK SVKSNLMOLFEESSGNTDIEGIDTNCYGTAAVFNAMVNESSWDGRYALVAGDI AIVATGNARPTGGVGAVALLIGNAPLIFDRGLRGTMOHAYDYKPDMLSEYPVVDG KLSTQCYLSALDRCYSVYRKRIQWQEGKDXDFTLNDGFMIFHSPYCKLVQKSLA RMFLNDFLNDQNRDNKNSISGLEAFGDVLEDTYFDRDVEKAFMKASSELFNQTKAS LIVSNQGNMVTSSVYGLASVLAQYSPQQLAGKRVGVFSYSGSLAATLYSLKVTODA TPGSALDKITASLCDLKSRLDSRTCVA PDVFAENMKLRDTHLANIYIPQCSIDSLFE GTWYLVRVDEKRRRTYARRPFTNDHSLDEGMGLVHSNTATEHLPSPAKKVPRLPATSA ESESAVISNGEH" 3208.3213 /note="putative" 3222 /note="putative"	
polYA_signal			
polYA_site			
ORIGIN			
Alignment Scores:			
Pred. No.:		3.34e-258	
Score:		2348.00	
Percent Similarity:		89.42%	
Best Local Similarity:		86.92%	
Query Match:		93.51%	
DB:		11 Gaps: 1	
US-10-622-516-2 (1-478) x AK045094 (1-3222)			
QY	1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProIyVaSpValGlyIleVal	20	
Db	62 ATGCTGGGTCACCTTCCTTGAATGCAGAGGCTTGCGCCAAAGATGTGGAAATCGTT	121	
QY	21 AlaleuGluIleTyRpheProSerGlnTyRValAspGlnAlaGluLeuGluTyRAsp	40	
Db	122 GCCCTTGAATCTACTTTCCTTCTCAATATGTCATCAAGCTGAGTTGAAAAATACGAT	181	
QY	41 GlyValAspAlaGlyLysTyRThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr	60	
Db	182 GGTGTAGATGCTGGAAGATATACCATCGGCTGGGCCACAGCCAGAGATGGCTTGCACG	241	
QY	61 AspArgGluAspIleAsnSerLeuCysMetThrValAlaGlnAsnLeuMetGluArgAsn	80	
Db	242 GATCGTGAAGACATCAACTCTTTGGCTGACTGTGTTCAAGAACTGATGGAGACAT	301	
QY	81 AsnLeuSerTyRAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys	100	
Db	302 AGCCTTCTCATGATTGATTCATTCGAGCTGTTGAGGAGCTCGGAATACAGATATA	361	
QY	101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle	120	
Db	362 TCGAATACAGTCAAGTCTAATTGATGACGCTGTTTGAGGAGCTCGGAATACAGATATA	421	
QY	121 GluGlyIleAspThrThrAsnAlaCysTyRGlYgLyThrAlaAlaValPheAsnAlaVal	140	
Db	422 GAAAGAAATAGATACCAACCATGCATGCTATGGGGGACACGCTGCAGTCTTCAATGCCGTG	481	
QY	141 AsnTrpIleGluSerSerSerTrpAsp-----	149	

Db 482 AACTGGGTCGAATCCAGCTCTGGGATGAGCATATGCTCTGTAGTTGCAGAGACATT 541
Qy 149 ----- 149
Db 542 GCTATATATGCCACAGAAATGCCAGACTACAGGTGAGTTGGAAGCTGTGGCCCTGCTA 601
Qy 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 602 ATTGGGCCAAACGCTCCTTAATTTTGTACCGAGGGCTCCGTGGACACACATGACGAT 661
Qy 159 AlaTyraSPheTyrlYsProaSPmetLeuSerGluTyProileValaSPGlyLysLeu 178
Db 662 GCCTATGACTTTACAAAGCCTGACATGCTCTCCGAGTACCCTGTGTCGACGAAAGCTC 721
Qy 179 SerileGlnCySTyrLeuSerAlaLeuAspArgCySTyrSerValTyrCySLysIle 198
Db 722 TCCATACAGTGCTACCTCAGCGCCCTGGACCGCTGCTATCTGTCTACCGCAAAAGATC 781
Qy 199 HisAlaGlnTrpGlnYsGlnGlyAsnaSPlysAspPheThrLeuAsnaSPheGlyPhe 218
Db 782 CGTGGCCAGTGGCAGAAAGAGGAAAGATTAAGATTTACCCTGAATGATTTGGCTTC 841
Qy 219 MetilePheHisSerProTyrcySLysLeuValGlnYsSerLeuAlaArgMetLeu 238
Db 842 ATGATCTTTCACTCACCATATTTGAACTGGTGACAGAAATCTTAGCTCGAGTTCCTG 901
Qy 239 AsnaSPheLeuAsnaSPGlnAsnaSPlysAsnSerileTySerGlyLeuGluAla 258
Db 902 AATGACTTTCTTAATGATCAAAACAGATAAAAACAGATTTACAGTGAGCTGGAAGCC 961
Qy 259 PheGlyAspValYsLeuGlnaSPThrTyrPheAspArgaSPValGluYsAlaPheMet 278
Db 962 TTTGGGAGCGTTAAATTAGAAAGTACTTACTTTGACAGAGATGTAGAAAAGCATTTATG 1021
Qy 279 LysAlaSerSerGluLeuPheSerGlnYsThrLysAlaSerLeuLeuValSerAnGln 298
Db 1022 AAGGCTAGTTCTGAGCTATTCAACAGAAAACAAAGGCGCTTTGCTGTCTTAATCAG 1081
Qy 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1082 AATGGAATATGTACACATCTCTGTATGGTTCCTGCTTCTGCTGCGACAGTAC 1141
Qy 319 SerProGlnGlnLeuAlaGlyLysArgileGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1142 TCACCTCAGCAGTTGGCAGGGAAGGGTGGAGTGTCTTTCACGTTCTGCGCTTGCT 1201
Qy 339 AlaThrLeuTyrSerLeuYsValThrGlnaSPAlaThrProGlySerAlaLeuAsPlys 358
Db 1202 GCCACGCTGTACTCCCTTAAGTCACACAAGACGCCACACAGAGATCTGCCCTGATAAA 1261
Qy 359 IleThrAlaSerLeuCyAspLeuYsSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1262 ATAAACAGCAAGTTATGTGACCTTAATCAAGGCTTGACTCGAAGCTGTGTGGCACCG 1321
Qy 379 AspValPheAlaGluAsnMetYsLeuArgGluAspThrHisIleValaSPntyrIle 398
Db 1322 GATGTCTTTGCTGAAAAACATGAAGCTCAGAGAGACACACATCACTTAGCCAACTATATT 1381
Qy 399 ProGlnGlySerileAspSerLeuPheGlnGlyThrTyPtyrLeuValArgValaSPGlu 418
Db 1382 CCCAGTGTCAATAGACTACTCTTTGAAGAACGTGTGATCTGCTCAGAGTGATGAA 1441
Qy 419 LysHisArgArgThrTyraAlaArgArgProThrProAsnaSPaSPThrLeuAspGlnGly 438
Db 1442 AAACACAGAAAGACTTACGCCCGCGCCCTTCACAAATGACACACAGTTTGATGAAGGA 1501
Qy 439 ValGlyLeuValHisSerAsnIleAlaThrGlnHisIleProSerProAlaYsIysVal 458
Db 1502 ATGGGGCTCGTCATAGTAACACAGCAACAGACATATTCAGCCCTGCTAAGAAAGTG 1561
Qy 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGlnHis 478

Db 1562 CCAAGACTCCCTGCAACCTCGGCCGAATCTGAATCAGCTGTCTATCAGTAACGGGGAGCAC 1621
RESULT 2
AK044835
LOCUS
DEFINITION
AK044835 3286 bp mRNA linear HTC 20-SEP-2003
Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
enriched library, clone:B13007E24 product:pre B-cell leukemia
transcription factor 1, full insert sequence.
ACCESSION
AK044835
VERSION
AK044835.1 GI:26336856
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636
REFERENCE
2
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
20499374
MEDLINE
11042159
PUBMED
REFERENCE
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL
20530913
MEDLINE
11076861
PUBMED
REFERENCE
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
JOURNAL
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
JOURNAL
6 (bases 1 to 3286)
MEDLINE
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,

COMMENT

Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. Tomohiro Kono (Department of Animal
Science, Tokyo University of Agriculture, 1737 Hnako Atsugi City,
Kanagawa Prefecture, Japan) whose assistance we gratefully
acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
Location/Qualifiers

FEATURES

Source

1..3286
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:B130007E24"
/db_xref="MGI:2410190"
/db_xref="taxon:10090"
/clone="B130007E24"
/tissue_type="parthenogenote"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9.5 days embryo"
123..1685
/note="unnamed protein product; pre B-cell leukemia
transcription factor 1 (MGD|MGI:97495, GB|NM_008783,
evidence: BLASTN, 99%, match=124)
putative"
/codon_start=1
/protein_id="BAC32112.1"
/db_xref="GI:26336857"
/translation="MPGSLPLNAEACWPKDVGIVALEIYFSPQYVDQALEKYGDVA
GKTTIGLGOARMGFCITREDINSLCTIVVQKLMERHSLSYDCIGRLVGTETIIDS
SVKSNLMOLFEEESGNTDIEGIDITNACYGTAVFNAVMVNESSMGRYALVAGDI
AIYATGNARPTGGVGAVALIIGPNAPLIFDRGLRGTHMQAYDFYKPDMLSEYVVDG
KLSTIQCYLSALDRCYSVYRKIRAOOWOKEGKDXDFTLNDGFMIFHSPYCKLVQKSLA
RMFLNDFLNDONRDKNSTIYSGLEAFGDYKLEDITYFBRDVEKAFMKASSELFNQKTKAS
LLVSNQNGNMWYTSVYGSVLASVLAQYSPQLAGKRYGVESYSGLAATLXSLKYODA
TPGSALDKITTASLCDLKRSLDRTCVAPDVFNAEMKLRBDTHLANIYIPQCSISLFE
GTWYLVAVDEKRRRTYARRPFTNDHSLDEGMGLVHSNTATEHIPPAKKVRLPATSA
ESESAVISNGEH"
3272..3277
/note="putative"
3286
/note="putative"

CDS

polyA_signal
polyA_site
ORIGIN
Alignment Scores:
Pred. No.:
Score: 3.45e-258 Length: 3286
Percent Similarity: 2348.00 Matches: 452
Best Local Similarity: 89.42% Conservative: 13
Query Match: 86.92% Mismatches: 13
DB: 93.51% Indels: 42
Gaps: 1

US-10-622-516-2 (1-478) x AK044835 (1-3286)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLyAspValGlyIleVal 20
Db 123 ATGCCTGGGTCACTTCCTTGAATGCAGAGGCTTGCTGGCCAAAGATGTGGAAATGCTT 182
QY 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db 183 GCCCTTGAATCTACTTCTCTCTCAATATGTGATCAAGCTGAGTTGGAATAATACGAT 242
QY 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 243 GGTGTAGATGCTGGAAGATATACCATCGGCTTGCGCCAGGCCAGGATGGCTTCTGCACG 302
QY 61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80

Db 303 GATCGTGAAGACATCACTCTTTGGCCTGACTGTGGTTCAGAACTGATGAGAGACAT 362
QY 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleLeuAspLys 100
Db 363 AGCCTTTCCTATGATTTGCATTTGGCGGCTAGAACTTGGAAACAGACAAATCATCGACAAA 422
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 423 TCGAATCAGTGAAGTCTAATTTGATGCAGCTGTTTGAGAGCTCTGGGAATACAGATATA 482
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 483 GAAGGAATAGATACAAACCAATGCATGCTATGGGGGCACAGCTGCAGTCTTCAATGCCGTG 542
QY 141 AsnTrpIleGluSerSerSerTrpAsp----- 149
Db 543 AACTGGGTGAAATCCAGCTCTTGGGATGACGATATGCTTGTAAGTTCAGAGACAT 602
QY 149 ----- 149
Db 603 GCTATATATGCCACAGAAATGCCAGACTACAGTGGAGTTGGAGCTGGCCCTGCTA 662
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 663 ATTGGGCCAAACGCTCCTCTTAATTTTGACCGAGGGCTCCGTGGACACACATGCAGCAT 722
QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db 723 GCCTATGACTTTTACAAAGCTGCATGCTCTCCGAGTACCCTGTGGTCGACGGAAGCTC 782
QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db 783 TCCATACAGTGTCACTACCTCAGCGCCCTGGACCGCTGCTATTTCTGTACCGCAAAAGATC 842
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 843 CGTGCCCACTGGCAGAAAGAGGAAAGATAAAGATTTACCCTGAATGATTTGGCTTC 902
QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 903 ATGATCTTTCACCTACCATATTGTAAACTGGTGCAGAAATCTCTAAGCTCGGATGTTCC 962
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 963 AATGACTTTCCTTAATGATCAAAACAGAGATAAAACAGTATTTCAGCTGGAAGGCC 1022
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 1023 TTTGGGAGACGTTAAATTGAAGATTACTTACTTGAACAGAGATGAGAAAAGGCATTTATG 1082
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 1083 AAGGCTAGTCTGAGCTATTCAACCAAGAAAAGGCGCTTTTGTCTTGAATCAG 1142
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 1143 AATGGAATATATATACACATCCTCTGTATGTTCCCTGGCTTCTGCTGGCACAGTAC 1202
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 1203 TCACCTCAGCAGTTGGCAGGGAAGAGGTGAGTGTCTCTTACGATTCTGGCTTGCT 1262
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 1263 GCCACGCTGTACTCCCTTAAAGTCACACACAGACGCCACACAGATCTGCCCTTGATAA 1322
QY 359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 1323 ATAACAGCAAGTTTATGTACCTTAAATCAAGGCTTGACTCGAGAACTTGTGTGCACCG 1382
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db 1383 GATGCTTTGCTGAAAACATGAAGCTCAGAGAGACACACATCATCTTACCACTATATT 1442

QY 399 ProGlnGlySerIleAspSerIleuPheGluGlyThrTrpTyrIleuValArgValaAspGlu 418
|||||
Db 1443 CCCCAAGTGTTCATAGACTCACTTTGGAAGAACGTGTATCTGTCAGAGTGTGATGAA 1502
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrIleuAspGluGly 438
|||||
Db 1503 AAACACAGAGGACTTACGCGCGCGCCCTTCACAAATGACACACAGTTTGATGAAGGA 1562
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysIleVal 458
:::|||||
Db 1563 ATGGGGCTCGTCATAGTACACAGCAACAGACATATTCAGCCCTGCTAAGAAAGTG 1622
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
|||||
Db 1623 CCAAGACTCCCTGCAACCTCGCGCGGAATCTGAATCAGCTGTATCAGTAACGGGAGCAC 1682
RESULT 3
AK031297
LOCUS 3466 bp mRNA linear HTC 18-SEP-2003
DEFINITION Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
enriched library, clone:6030403N11 product:pre B-cell leukemia
transcription factor 1, full insert sequence.
ACCESSION AK031297 GI:26327188
VERSION AK031297.1 GI:26327188
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
1 High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE
AUTHORS Adachi, J., Aizawa, K., Akiyama, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hasegaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
FEATURES
Source location/Qualifiers
1. .3466
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:6030403N11"
/db_xref="MG1:2394508"
/db_xref="taxon:10090"
/clone="6030403N11"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
122. .1684
/note="unnamed protein product; pre B-cell leukemia
transcription factor 1 (MGD|MG1:97495, GB|NM_008783,
evidence: BLASTN, 99%, match=124)
putative"
/codon_start=1
/protein_id="BAC27338.1"
/db_xref="GI:26327189"
/translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYGVDVA
GKYTIGLGQARMGCTDRDINSLCTTVQKLMERHSLSYDCIGRLVETGTEIIDSKS
SVKSNLMOLFEEBSGNTDIEDIDTNACYGGTAVERNVMWVSSSWDGRYALVAVGDI
AIYATGNARPTGGVGAVALLIGPNAPLIFDRGLRGTHMOHAYDFYKPKMLSEYVVDG
KLISQCYLSALDRCSYVRKKIRAOQKEGKDDFTLNDFGMI FHSPYCKLVQKSLA
RMFLNDFLNDQNDKNSIYSGLEAFGVDKLEDITYFDRDVEKAFMKASSELFNQTKAS
LLVSNQNGMWTSSVYSGSLASVLAQVSPQLAGRIVGVFSYSGSLAATLVSLKTQDA
TPGSALDIKTASLCDLKSRLDSRTCVAPDVFAENMKLRDTHLANYIPQCSIDSLFE
GTWLVLRVDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATEHIPSAPKVPRLPATSA
ESESAVISNGEH"
3447. .3452
/note="putative"
3466
/note="putative"
ORIGIN
polyA_signal
polyA_site
Alignment Scores:
Pred. No.: 3.77e-258 Length: 3466
Score: 2348.00 Matches: 452
Percent Similarity: 89.42% Conservative: 13
Best Local Similarity: 86.92% Mismatches: 13
Query Match: 93.51% Indels: 42
DB: 11 Gaps: 1
US-10-622-516-2 (1-478) x AK031297 (1-3466)
QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
|||||


```
Db      122 ATGCGTGGGTCACTTCCTTGAATGCAGAGGCTTGCTGGCCAAAGAATGTGGGAATCGTT 181
QY      21  AAlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db      182 GCCCTTGAATATCACTTCTCCTTCATATATGTCATCAAGCTGAGTTGAAAAATACGAT 241
QY      41  GJyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db      242 GGTGTAGATGCTGGAAAAGTATACCATCGGCTGGCCAGGCCAGGATGGCTTGACAG 301
QY      61  AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn 80
Db      302 GATCGTGAAAGACATCAACTCTCTTGCTGCTGACTGTGTTCAAAAACTGATGAGAGACAT 361
QY      81  AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db      362 AGCCTTTCCTATGATTCGATTTGGCGGCTAGAGTTGGAACAGACATCATGCACAAA 421
QY      101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db      422 TCGAAATCAGTGAAGTCTAATTGATGCAGCTGTTGAGAGTCTGGGAATACGATATA 481
QY      121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db      482 GAAGAATAGATCAACCAATGCTATGCGGCGACAGCTGCAGTCTCAATGCCGTG 541
QY      141 AsnTriPleGluSerSerSerTyrAsp----- 149
Db      542 AACTGGGTGGAATCCAGCTCTGGGATGAGCAGATATGCTGTGTAATTGCAGAGACATT 601
QY      149 ----- 149
Db      602 GCTATATATGCCACAGAAATGCCAGACTACAGGTGAGTTGGAGCTGTGCCCTGCTA 661
QY      150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db      662 ATTGGGCCAAACGCTCCTTAATTTTTTGACCGAGGGCTCCGTGGGAACACACATGCAGCAT 721
QY      159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db      722 GCCTATGACTTTTACAAGCCGTGCATGCTCTCCAGTACCCTGTGTCGACGCGAAAAGCTC 781
QY      179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db      782 TCCATACAGTGCCTACCTCAGCGGCTGCAGCCGCTATTCGTCTACCGCAAAAAGATC 841
QY      199 HisAlaGlnTyrGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db      842 CGTGCCCAAGTGCAGAAAGAGGAGGAAAGATTAATTTACCCTGAATGATTTGGCTTC 901
QY      219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db      902 ATGATCTTCACTCACCATATTTGTAACCTGTGCAGAAATCTCTAGCTCGGATGTTCCCTG 961
QY      239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db      962 AATGACTTCTTAATGATCAAAAACAGATATAAAACAGTATTTCAGTGGACTGGAAGCC 1021
QY      259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db      1022 TTTGGGGACGTTAAATTAGAGATACTTACTTTGACAGAGATGTAGAAAAGGCATTATG 1081
QY      279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db      1082 AAGGCTAGTCTGAGCTATTCAACAGAAAACAAAGCGCTTTGCTGTGTCTAATCAG 1141
QY      299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db      1142 AATGGAATATGTACACATCCTCTGTCTATGTTCCCTGGCTTCTGCTGCACAGTAC 1201
QY      319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db      1202 TCACCTCAGCAGGTGGCAGGGAAGAGGTTGAGTGTCTCTTAACGGTCTGGCTTGCT 1261
```

```
QY      339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db      1262 GCCACGCTGACTCCCTTAAAGTCAACAAAGCGCCACACAGGATCTGCCCTTGATAAA 1321
QY      359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db      1322 ATAAACAGCAAGTTTATGTGACCTTAAATCAAGGCTTGACTCGAGAACTTGTGTGGCACCG 1381
QY      379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle 398
Db      1382 GATGCTTGTGGAACAAACATGAAGCTCAGAGAGGACACACATCACTTAGCCAACTATATT 1441
QY      399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTyrTyrLeuValArgValAspGlu 418
Db      1442 CCCCAAGTTCATAGACTCACTCTTTGAAGAAACGTGTATCTGTCAGAGTGGATGAA 1501
QY      419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db      1502 AAACACAGAAAGACTTACCGCCCGCGCCCTTCACAAATGACACACAGTTTGAAGAAGA 1561
QY      439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db      1562 ATGGGGCTCGTCATAGTAAACACAGCAACAGACATATTCGAAGCCCTGTAAGAAAGTG 1621
QY      459 ProArgLeuProAlaThrAlaAlaGluProGluAlaValAlaIleSerAsnGlyGluHis 478
Db      1622 CCAAGACTCCCTGCAACCTCGGCCGAATTCGAATCAGCTGTCAATCAATAACGGGGAAGCAC 1681

RESULT 4
AY416670
LOCUS      1527 bp      DNA      linear      GSS 17-DEC-2003
DEFINITION Homo sapiens HMGS2 gene, VIRtual TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY416670
VERSION     AY416670.1  GI:39772630
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 1527)
            Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
            Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
JOURNAL     Science 302 (5652), 1960-1963 (2003)
PUBMED      14671302
REFERENCES   2 (bases 1 to 1527)
            Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
            Direct Submission
            Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT      This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES
            source
                1..1527
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                <1..>1527
                /gene="HMGS2"
                /locus_tag="HCM5968"
ORIGIN
Alignment Scores:      1.36e-162      Length:      1527
Pred. No.:             1513.00      Matches:      282
Score:
```

Percent Similarity: 75.16% Conservative: 69
Best Local Similarity: 60.39% Mismatches: 74
Query Match: 60.25% Indels: 42
DB: 29 Gaps: 1

US-10-622-516-2 (1-478) x AY416670 (1-1527)

QY 4 SerLeuProLeuAsnAlaGluAlaCysTrpProLyAspValGlyIleValAlaLeuGlu 23
Db 121 GCTGTCCCCCTGGCCAAACAGATACTTGGCCAAAGACGTGGGCATCTGTGCCCTGGAG 180
QY 24 IleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAspGlyValAsp 43
Db 181 GTCTACTTCCAGCCCAATATGTGACCAAACTGACCTGGAAGATATAACAATGTGAA 240
QY 44 AlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThrAspArgGlu 63
Db 241 GCAGGAAAGTATACAGTGGGCTTGGCCAGACCCTATGGGCTTCTGCTCAGTCCAGAG 300
QY 64 AspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsnAsnLeuSer 83
Db 301 GACATCAACTCCCTGTGCTGCGTGAACGGTGTGCAACGGCTGATGAGCGCATACAGTCCCA 360
QY 84 TyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleLeAspLysSerLysSer 103
Db 361 TGGGACTCTGTGGGAGGCTGGAGTAGGCACTGAGACCATCATTGACAAGTCCAAAGCT 420
QY 104 ValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIleGluGlyIle 123
Db 421 GTCAAAACAGTGCTCATGGAACTCTCCAGGATTCAGGCAATACGTATATTGAGGGCATA 480
QY 124 AspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaValAsnTrpIle 143
Db 481 GATACCAACCAATGCGCTGCTACGGTGTACTGCTCCCTCTTCAATGTGCCAACTGATG 540
QY 144 GluSerSerSerTrpAsp----- 149
Db 541 GAGTCCAGTTCCTGGGATGGTCGTTATGCCATGTGTGTGTGGAGACATTGCCGCTTAT 600
QY 149 ----- 149
Db 601 CCCAGTGTATATGCTCGTCCACAGGTGGGCGCGAGCTGTGGCTATGCTGATTGGGCCC 660
QY 150 -----GlyLeuArgGlyThrHisMetGlnHisAlaTyrAsp 161
Db 661 AAGGCCCTCTGGCCCTGGAGCGAGGCTGAGGGGAACCCATATGAGAATGTGTATGAC 720
QY 162 PheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIleGln 181
Db 721 TTCTACAAACCAATTTGGCCCTGGAGTAGCCCAATAGTAGGAAAGCTTTCATCCAG 780
QY 182 CysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysIleHisAlaGln 201
Db 781 TGCTACTTGGGGCGCTTGATCGATGTTATACACATCATACCGTAAATAATCCAGAATCAG 840
QY 202 TrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIlePhe 221
Db 841 TGGAAAGCAAGCTGGCAGCGATCGACCCCTTCAACCCTTGACGATTTTACAGTACATGATCTT 900
QY 222 HisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAspPhe 241
Db 901 CATACACCCCTTTGGCAAGATGTCCAGAAGTCTGTGGCTCGCTGATGTTCAATGACTTC 960
QY 242 LeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGlyAsp 261
Db 961 CTGTCAAGCCAGCAGTGACACACAAACCAAGCTTATATAAGGGGCTGAGGCTTTCGGGGG 1020
QY 262 ValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAlaSer 281
Db 1021 CTAAGCTGGAAGACACCTACACCAACAAGAGACCTGATAAAGCACTTCTAAAGGCTCT 1080
QY 282 SerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGlyAsn 301

Db 1081 CAGGACATGTTGACAGAAACCAAGGCTTCCCTTTTACCTCTCCACTCACATGGGAAC 1140
QY 302 MetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerProGln 321
Db 1141 ATGTACACCTCATCCCTGTACGGGTGCTGCGCTTCTGTCTGTCTGCCACCACTCTGCCCA 1200
QY 322 GlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThrLeu 341
Db 1201 GAACTGGCTGGCTCCAGGATGTGCTTCTCTTATGGCTGTGTTAGCAGCAAGTTTC 1260
QY 342 TyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThrAla 361
Db 1261 TTTTCATTTCCAGATATCCAGGATGCTGCTCCAGGCTCTCCCTGACACAGTTGTTCC 1320
QY 362 SerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAspValPhe 381
Db 1321 AGCACATCAGACCTGCCAAACGCCCTTAGCCTCCCGAAAGTGTGTCTCTCAGAGATTTC 1380
QY 382 AlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIleProGlnGly 401
Db 1381 ACAGAAATATGAAACCAAGAGACCAATTCTACCATAGGTGAATTCTCCCACTGTGT 1440
QY 402 SerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGluLysHisArg 421
Db 1441 GACACAACACGCTTTTCCAGGTACTTGTGTAACCTGAGCGAGTGACGACGACGATCGC 1500
QY 422 ArgThrTyrAlaArgArgPro 428
Db 1501 CGAAAGTATGCCCGGCGTCCC 1521

RESULT 5
AY416672 1527 bp DNA linear GSS 17-DEC-2003
LOCUS AY416672
DEFINITION Mus musculus HMGCS2 gene, VIRUTAL TRANSCRIPT, partial sequence,
Genomic survey sequence.
ACCESSION AY416672
VERSION AY416672.1 GI:39772632
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1527)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1527)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES
source
1..1527
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1527
/gene="HMGCS2"
/locus_tag="HCM5968"

ORIGIN
Alignment Scores: 2.74e-160 Length: 1527
Pred. No.:

Score: 1493.00 Matches: 276
Percent Similarity: 75.27% Conservative: 74
Best Local Similarity: 59.35% Mismatches: 73
Query Match: 59.46% Indels: 42
DB: 29 Gaps: 1
US-10-622-516-2 (1-478) x AY416672 (1-1527)

QY 6 ProLeuAsnAlaGluAlaCySerTrpProLysAspValGlyIleValAlaLeuGluIleTyr 25
Db 127 CCCCTGGCCAAACAGATACATGGCCAAAAGATGTGGCATCCTTGCCCTGGAGGTCTAT 186
QY 26 PheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAspGlyValAspAlaGly 45
Db 187 TTTCAGCCCAATATGTGACCAAACTGACCTGAAAAAGTTCAACAATGTGAAGCAGGG 246
QY 46 LysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThrAspArgGluAspIle 65
Db 247 AAGTATACAGTGGGCTTGCGCCAGACCCTATGGGCTTCTGTTCAGTCCAGGAGACATC 306
QY 66 AsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsnAsnLeuSerTyrAsp 85
Db 307 AACTCCCTGTGCTGACAGTGTACAGAGGCTGATGGAACGCACAAAGCTGCCGTGGAT 366
QY 86 CysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLysSerLysSerValLys 105
Db 367 GCTGTGGCGCTGTGGAAGTGGGACCGAGACCATCATTGACAAGTCCAAAGGCTGTCAA 426
QY 106 ThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIleGluGlyIleAspThr 125
Db 427 ACAGTGTCTCATGGAACCTGTTCCAGGATTCAGGCAACACTGACATCGAGGCAATAGATACC 486
QY 126 ThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaValAsnTrpIleGluSer 145
Db 487 ACCAACGCGCTGTATGCGGCGACAGCCTCCCTTCAATGCTGCCCACTGAGTGAAGTCC 546
QY 146 SerSerTrpAsp----- 149
Db 547 AGCTACTGGAGTGTGCTATGCCCTGTGTGTGTGACATTGCACTACCCGAGT 606
QY 149 ----- 149
Db 607 GGTAACGCCCCGCCACAGGTGTGCTGGGCTGTGGCAATGCTGATCGGGCCCAAGGCC 666
QY 150 -----GlyLeuArgGlyThrHisMetGlnHisAlaTyrAspPheTyr 163
Db 667 CCTCTGTCCTCGAGCAAGGGCTGAGGGGAACCTCACATGAGAAACGCCGTAACGACTTCTAC 726
QY 164 LysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIleGlnCysTyr 183
Db 727 AAACCAAACTGGCCTCAGAGTATCCACTGTGTGATGGGAAGCTGTCTATCCAGTGTCTAC 786
QY 184 LeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAlaGlnTrpGln 203
Db 787 CTGCGGGGCTTGATCGATGCTATGCAGCCTACCGCAAGAGATCCAGATCAGTGAAG 846
QY 204 LysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIlePheHisSer 223
Db 847 CAAGCTGGAACAAACACGACCTTCAACCTCGATGATGTGCAGTATATGATCTTCCACACA 906
QY 224 ProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsn 243
Db 907 CCCTTTTGCAAGATGTCACAGAAATCCCTGGCTGCTGATGTTCATGACTTCTCTGCA 966
QY 244 AspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGlyAspValLys 263
Db 967 TCCAGCAGTGCACAAACAGAACACTTATACAAGGCCCTAGAGGCCCTTCAGGGGCTTAAAG 1026
QY 264 LeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAlaSerSerGlu 283
Db 1027 CTGGAAGAAACCTACACCAACAGGATGAGCAAGGCTCTTTGAAGGCTCCTCGTGAAC 1086
QY 284 LeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyr 303

Db 1087 ATGTTCAACCGAAGAACCAAGGCTCCCTTTACCTCTCCACAAATATATGGAACTGTAC 1146
QY 304 ThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerProGlnGlnLeu 323
Db 1147 ACCCTCTCCCTTAATGCTGTGCTGGCCTCCTCACTTCTCTCTCACCACCTCTGCCCAAGATTG 1206
QY 324 AlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThrLeuTyrSer 343
Db 1207 GCTGGCTCCAGGATTGAGGCTTCTCTACGGCTCAGGCTTAGCAGCAAGTTCTTTTCA 1266
QY 344 LeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThrAlaSerLeu 363
Db 1267 TTCCGAGTGTCCAAAGGATGCTTCCCAAGTTCCCTGGAGAAACTGTGTCTAGTGTG 1326
QY 364 CysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaAlaProAspValPheAlaGlu 383
Db 1327 TCAGATCTGCCCAACGCTTAAGCTCCCGAGACGCATGTCCCTGAGGAATTTCAGAA 1386
QY 384 AsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIleProGlnGlySerIle 403
Db 1387 ATAATGAACCAAGAGAGCAATTTCACCAAGGTGAACCTTCTCTCCACCTGAGACACA 1446
QY 404 AspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGluLysHisArgArgThr 423
Db 1447 AGCAATCTCTTCCAGGTACTGTGTACCTTGAACGAGTGATGATGATCATCGCAGGAAG 1506
QY 424 TyrAlaArgArgPro 428
Db 1507 TATGCCCGGTGTCCC 1521
RESULT 6
AK004865 3287 bp mRNA linear HTC 20-SEP-2003
LOCUS
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300002P16
product:3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2, full insert sequence.
ACCESSION AK004865
VERSION AK004865.1 GI:12836370
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 3287)

AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAGATCCAGAGCTCAATTAATTAACCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

Source location/Qualifiers

1..3287

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="PANTOM:DB:1300002P16"

/db_xref="MGI:19072269"

/db_xref="taxon:10090"

/clone="1300002P16"

/sex="male"

/tissue_type="liver"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

52..1578

/note="unnamed protein product; 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (MGD|MGI:101939)"

CDS

/codon_start=1

/protein_id="BAB23626.1"

/db_xref="GI:12836371"

/translation="MQRLLAPARRVLQVKRAMQETSLTPAHLLSAQQRFSITPPAPLAKDTWPKDVGILALEVYFPAQVVDQTDLEKFNVEAGKYTVGLGQTRMGFCSQVEDINSICLTVQRLMERTKLPMDAVGRLEVGTEITIDSKAVKTVLMELFQDSGNTDIEGIDTNACYGCTASLFNANMNMSSSYWDGRYALVVGEDIAVYPSGNARPTGAGAVAMLIGPKAPLVLEOGLRGTHMENAYDYFKPNLASEYPLVDGKLISQCYLRALDRCYAAVRKKIQNQWKQAGNNQPFITLDVQYMI FHTPFCKMVQKSLARLMFNDFLSSSSDKQNNLYKGLAFAFRGLKLEETYNKDVKALLKASLDMFNQKTRASLYLSTNNGNMYTSLYGLIAS

LSHSAQELACSRIGAFSYGSLAASFPRVSKDASPGSPLEKLVSSVSDLPKRLD
SRRMSPDEFTEIMNQREQFYHKVNFSPPGDTSNLFPGTWYLERVDEMRKXARCPV

polYA_signal 3266..3271
polYA_site 3287
/note="putative"
/note="putative"

ORIGIN

Alignment Scores:

Pred. No.:	9.84e-160	Length:	3287
Score:	1493.00	Matches:	276
Percent Similarity:	75.27%	Conservative:	74
Best Local Similarity:	59.35%	Mismatches:	73
Query Match:	59.46%	Indels:	42
DB:	11	Gaps:	1

US-10-622-516-2 (1-478) x AK004865 (1-3287)

QY 6 ProLeuAsnAlaGluAlaCysTrpProLyAspValGlyIleValAlaLeuGluIleTyr 25

DB 178 CCCCTGGCCAAACAGATACATGCCCCAAAGATGTGGCATCTTGCCCTGAGGTCTAT 237

QY 26 PheProSerGlnTyrValAspGlnAlaGluLeuGluIleTyrAspGlyValAspAlaGly 45

DB 238 TTTCACGCCCAATATATGTGACCAACTGACCTGGAAGTTCAACAATGTGGAAGCAGGG 297

QY 46 LysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThrAspArgGluAspIle 65

DB 298 AAGTATACAGTGGGCTTGGCCAGACCCGATATGGGCTTCTGTCACTCAGTCAGAGGACATC 357

QY 66 AsnSerLeuCysMetThrValAlaGlnAsnLeuMetGluArgAsnAsnLeuSerTyrAsp 85

DB 358 AACTCCCTGTGCTGACAGTGTACAGAGCTGATGGAACGACAAAGCTGCCGTGGAT 417

QY 86 CysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLysSerLysSerValLys 105

DB 418 GCTGTGGCGCGTCTGGAAGTGGGCAACGAGACCATCATGACAACTCCAAGCTGTCAA 477

QY 106 ThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIleGluGlyIleAspThr 125

DB 478 ACAGTCTCATGGAAGTGTTCAGGATTCAAGCAACACTGACATCGAGGGCATAGATACC 537

QY 126 ThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaValAsnTrpIleGluSer 145

DB 538 ACCAACGCTGTATAGCGGACAGCCTCCTCTTAATGCTGCCAATGATGATGAGTGC 597

QY 146 SerSerTrpAsp----- 149

DB 598 AGCTACTGGATGTGCTATGCCCTGTGTGTGTGATGACATTGCAGTCTACCCGAGT 657

QY 149 ----- 149

DB 658 GGTAACGCCGCCCCACAGAGTGCTGGGGCTGTGGCAATGCTGATCGGGCCAAAGGCC 717

QY 150 -----GlyLeuAArgGlyThrHisMetGlnHisAlaTyrAspPheTyr 163

DB 718 CCTGTGCTCCTCGAAGCAGGGCTGAGGGAACCTCAATGAGAAACCGTACGACTTCTAC 777

QY 164 LysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIleGlnCysTyr 183

DB 778 AAACCAAACTTGGCCTCAGAGTATCCACTGTGTGATGGGAAGCTGTCTATCCAGTGTAC 837

QY 184 LeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAlaGlnTrpGln 203

DB 838 CTGGCGGCGCTTGATCGATGCTATGACACCTACCGCAAGAGATCCAGAATCAGTGAAG 897

QY 204 LysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIlePheHisSer 223

DB 898 CAAGCTGGAACAACACAGCCCTTCAACCTCGATGATGTGCAATATATGATCTCCACACA 957

QY 224 ProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsn 243

Db 958 CCCTTTTGCAAGATGCTCCAGAAATCCCTGGCTCGGTGTGATGTCAATGACTTCTGTCA 1017

QY 244 AspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGlyAspValLys 263
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1018 TCCAGCAGTGCACAACAGACAACACTTATACAAGGGCTTAGAGGCTTCAGGGGCTAAAG 1077

QY 264 LeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAlaSerSerGlu 283
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1078 CTGGAAGAAACCTACACCAACAAGGATGTAGACAAGGCTCTTTGAAGGCCCTCCCTGGAC 1137

QY 284 LeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyr 303
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1138 ATGTTCAACCAAGAGCCCAAGGCCCTTACCTCTCCACAATAATGGAACATGTAC 1197

QY 304 ThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerProGlnGlnLeu 323
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1198 ACCTCTTCCCTCTATGGCTGCTCGGCTCCTCCTCTCTCTCACCACCTGCCCCAAGAATTG 1257

QY 324 AlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThrLeuTyrSer 343
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1258 GCTGGCTCCAGGATGGAGCCTTCTCTACGGCTCAGGCTTAGCAGCAAGTTCTTTTCA 1317

QY 344 LeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThrAlaSerLeu 363
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1318 TTCCGAGTGTCCAAGGATGCTTCCCAAGTTCCCCCTTGAGAAACTGTGTCTAGTGTG 1377

QY 364 CysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAspValPheAlaGlu 383
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1378 TCAGATCTGCCCAACGCTCTAGACTCCCGAGAGCGCATGTCCCTGAGGAATTCAAGAA 1437

QY 384 AsnMetLysLeuArgGluAspThrHisLeuValAsnTyrIleProGlnGlySerIle 403
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1438 ATAATGAACCAAGAGAGCAATTTTACACAGAAGTGAATCTCTCCACCTGAGACACA 1497

QY 404 AspSerLeuPheGlnGlyThrTyrTyrLeuValArgValAspGlnLysHisArgArgThr 423
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1498 AGCAATCTCTTCCCAAGTACTTGTAACCTTGAAAGATGATGATGCATCGCAGGAAG 1557

QY 424 TyrAlaArgArgPro 428
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1558 TATGCCCGGTGCC 1572

RESULT 7

AK004902 2524 bp mRNA linear HTC 20-SEP-2003

LOCUS

DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300004J23

product:3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2, full insert sequence.

AK004902

VERSION AK004902.1 GI:12836438

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS 1 Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE

AUTHORS 2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE

AUTHORS 3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,

Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913

11076861

4

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 2524)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoaka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAGATCCAGAGCTCAATTAATTAAATTAACCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

source

1..2524

location/qualifiers

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:1300004J23"

/db_xref="MGI:1910606"

/db_xref="taxon:10090"

/clone="1300004J23"

/sex="male"

/tissue_type="liver"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

50..1576

/note="unnamed protein product; 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (MGD|MGI:101939)

```
putative"
/codon_start=1
/protein_id="BAB23657.1"
/db_xref="GI:12836439"
/translation="MQRLAPARRVLQVKRAMQETSLTPAHLLSAQORFSTIPAPL
AKDTWPKDVGILALEVYFPAQYVDQTDLEKFNVEAGKYTVGLGQTRMGSCVQEDI
NSLCLTVQRLMERTKLPMDAVGRLEVGTEITIDSKAVKTVLMELFQDSGNTDIEGI
DTNACVGGTASLFRNANWMESSYWDGRYALVVGDIAYVPSGNARPTGAGAVAML I
GPKAPLVLEOGLRGTHMENAYDFKPNLASEVPLVDGKLSIOCYLRALDRCYAAVRKK
IQNQWKQAGNNQPTLDDVQYMI FHTPFCRMVQKSLARLMFNDPLSSSDKQNNLYKG
LEAFRGLKLEETYNKDQVDKALLKASLDMFNQKESLSTNNGMNMTSLYGCLAS
LISHSAQELAGSRIGAFSYSGLAASFPSFVSKDASPGSPLEKLVSSVSDLPKRLD
SRRMSPEEFTEIMNQREQFYHKVNPSPGDTSNLFPGTWYLERVDEMHRKRYARCPV
"
polya_signal      2503. .2508
                   /note="putative"
polya_site        2524
                   /note="putative"
ORIGIN
Alignment Scores:
Pred. No.:        1.08e-159      Length:      2524
Score:            1491.00         Matches:     275
Percent Similarity: 75.27%        Conservative: 75
Best Local Similarity: 59.14%      Mismatches:  73
Query Match:      59.38%          Indels:       42
DB:               11              Gaps:         1
US-10-622-516-2 (1-478) x AK004902 (1-2524)
QY      6 ProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleValAlaLeuGluIleTyr 25
Db      176 CCCCTGGCCAAACAGATACATGCGCCAAAGATGTGGGCATCTTGCCCTGAGGCTTAT 235
QY      26 PheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAspGlyValAspAlaGly 45
Db      236 TTTCAGCCCAATATGTGACCAACTGACCTGMAAAGTTCAACAATGTGAAGCAGGG 295
QY      46 LysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThrAspArgGluAspIle 65
Db      296 AAGTATACAGTGGCTTGGGCCAGACCCGTATGGCTTCTGTTCAGTCCAGAGAGACATC 355
QY      66 AsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsnAsnLeuSerTyrAsp 85
Db      356 AACTCCCTGTGCTGACAGTGTGTCAGAGGCTGATGGAACGCACAAAGCTGCCGTGGAT 415
QY      86 CysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLysSerLysSerValLys 105
Db      416 GCTGTGGCCGCTCTGGAAGTGGGCACCGAGACCATCATTTGACAAAGCTGTCTAAA 475
QY      106 ThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIleGluGlyIleAspThr 125
Db      476 ACAGTGTCTCATGGAAGCTGTTCCAGGATTCAAGCAACACATGCATCGAGGGCATAGATAC 535
QY      126 ThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaValAsnTrpIleGluSer 145
Db      536 ACCAAGCCTGTATTATGCGGCAAGCCTCCCTCTTCAATGCTGCCAACTGATGAGTCC 595
QY      146 SerSerTrpAsp----- 149
Db      596 AGCTACTGGGATGTGCTATGCCCTGTGTGTCTGTGTGACATTGCAGTCTACCCGAGT 655
QY      149 ----- 149
Db      656 GGTAACGCGCGCCACACAGGTGCTGGGGCTGTGGCAATGCTGATCGGGCCCAAGGCC 715
QY      150 -----GlyLeuArgGlyThrHisMetGlnHisAlaTyrAspPheTyr 163
Db      716 CTTCTGGTCTCGAGCAAGGGCTGAGGGGAATCACAATGAGAAACGCGTACGACTTCTAC 775
QY      164 LysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIleGlnCysTyr 183
Db      776 AAACCAAACTTGGCCTCAGAGTATCCACTGTGTGATGGGAAGCTGTCTATCCAGTGTCTAC 835
```

```
QY      184 LeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAlaGlnTrpGln 203
Db      836 CTGCGGGCCCTTGATCGATGCTATGCAAGCCTACCCGCAAGATCCAGAATCAGTGAAG 895
QY      204 LysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIlePheHisSer 223
Db      896 CAAGCTGGAACAACACAGCCTTTCAACCTCGATGATGTGAGTATATGATCTTCCACACA 955
QY      224 ProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsn 243
Db      956 CCTTTTGCAGATGTGTCAGAAATCCCTGGCTCGGTGATGTTCAATGACTTCCGTCA 1015
QY      244 AspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGlyAspValLys 263
Db      1016 TCCAGCAGTGACAAACAGAACACTTATACAGGGCCTGAGGCCTTCAAGGCTTAAG 1075
QY      264 LeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAlaSerSerGlu 283
Db      1076 CTGGAAGAAACCTACACCAACAAGATGTAGACAAAGCTCTTTGAAGGCCTCCCTGAC 1135
QY      284 LeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyr 303
Db      1136 ATGTTCAACCAAGAACCAAGCCTCCTTACCTCTCCACAATATATGGGAACATGTAC 1195
QY      304 ThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerProGlnGlnLeu 323
Db      1196 ACCTCTTCCCTATAGCTGCTGCTGCTCCTCACTTATCTTCAACCACTCTGCCAAGAAATG 1255
QY      324 AlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaIleThrLeuTyrSer 343
Db      1256 GCTGGCTCCAGATTGAGCCTTCTCCTACGCTCAGGCTTAGCAAGATTCTTTTCA 1315
QY      344 LeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThrAlaSerLeu 363
Db      1316 TTCCGAGTGTCCAAGATGCTTCCCAAGTTCCCCCTGGAGAAACTGCTGTAGTGTG 1375
QY      364 CysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAspValPheAlaGlu 383
Db      1376 TCAGATCTGCCCAACGTTTAGACTCCCGGAGACGCAATGTCCTGAGGAATTCAAGAA 1435
QY      384 AsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIleProGlnGlySerIle 403
Db      1436 ATTAATGAACCAAGAGAGCAATTTCACACAAAGTGAACCTTCTCCACCTGAGACACA 1495
QY      404 AspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGluLysHisArgArgThr 423
Db      1496 AGCAATCTCTTCCAGGTACTTGTAACCTTGAACGAGTGAATGATGCATCGCAGGAAG 1555
QY      424 TyrAlaArgArgPro 428
Db      1556 TATGCCCGGTGTCCC 1570
RESULT 8
BX441322      1083 bp      mRNA      linear      EST 15-MAY-2003
LOCUS      BX441322      Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION      CS0DF016YJ04 5-PRIME, mRNA sequence.
ACCESSION      BX441322
VERSION      BX441322.1      GI:30789948
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      1 (bases 1 to 1083)
JOURNAL      Full-length cDNA libraries and normalization
COMMENT      Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
```


Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3098.r For more information about this cluster, see

http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODF016DE02QP1&cluster=3098.r. Contact : Peng Liang Email : liang@life.techn.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope Sequence ID : CSODF016DE02QP1.

FEATURES

source

1..1083
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODF016YJ04"
/issue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Alignment Scores:

Pred. No.: 9.35e-148 Length: 1083
Score: 1382.00 Matches: 263
Percent Similarity: 86.23% Conservative: 0
Best Local Similarity: 86.23% Mismatches: 0
Query Match: 55.04% Indels: 42
DB: 13 Gaps: 1

US-10-622-516-2 (1-478) x BX441322 (1-1083)

QY 1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProIysAspValGlyIleVal 20
Db 168 ATGCCTGGMTCACTTCCTTGAATGAGAAGCTGCTGGCCAAAGATGTGGGAATTGTT 227
QY 21 AlaLeuGluIleTyrrPheProSerGlnTyrrValAspGlnAlaGluLeuGluIysTyrrAsp 40
Db 228 GCCCTTGAGATCTATTTTCTTCTCAATATGTTGATCAAGCAGAGTTGGAATAATATGAT 287
QY 41 GlyValAlaSpAlaGlyLysTyrrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db 288 GGTGTAGATGCTGGAAGAATATACCATTTGGCTGGGCCCAAGGCATGGGCTTCTGCACA 347
QY 61 AspArgGluAspIleAsnSerLeuCyseMetThrValValGlnAsnLeuMetGluArgAsn 80
Db 348 GATAGAGAAGATATTAACCTCTCTTGATGACAGCTGCTTCAAGATCTTATGGAGAGAAAT 407
QY 81 AsnLeuSerTyrrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db 408 AACCTTCTCTATGATTGCAATTTGATGACAGCTGTTGAAGAGTCTGGGAATACAGATATA 467
QY 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 468 TCAAAGTCTGTGAAGACTAATTTGATGACAGCTGTTGAAGAGTCTGGGAATACAGATATA 527
QY 121 GluGlyIleAspThrThrAsnAlaCysTyrrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db 528 GAAGGAATGCACACAACTAATGCAATGATGAGGACACAGCTGCTGTCTCATGCTGTT 587
QY 141 AsnTrpIleGluSerSerSerTyrrAsp----- 149
Db 588 AACTGGATTGAGTCCAGCTCTTGGGATGACCGGTATGCCCTGGTAGTTGCAGAGATATT 647
QY 149 ----- 149
Db 648 GCTGTATATGCCACAGAAATGCTAGACCTACAGGTGAGTTGAGACAGTACGCTCTGCTA 707
QY 150 -----GlyLeuArgGlyThrHisMetGlnHis 158
Db 708 ATGGGCCCAATGCTCCTTAAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT 767

QY 159 AlaTyrrAspPheTyrrLysProAspMetLeuSerGluTyrrProIleValAlaSpGlyLysLeu 178
Db 768 GCCTATGATTTTTCACAGCCCTGATATGCTATCTGAATATCCTATAGTAGATGAAAACTC 827
QY 179 SerIleGlnCysTyrrLeuSerAlaLeuAspArgCysTyrrSerValTyrrCysLysLysIle 198
Db 828 TCATACAGTGTCTACCTCAGTGCAATTAGACCGCTGCTACTCTGTCTACTGCMAAAGATC 887
QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db 888 CATGCCAGTGGCAGAAAGAGGAAATGATTAAGATTTTACCCTTGAATGATTTGGCTTC 947
QY 219 MetIlePheHisSerProTyrrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db 948 ATGATCTTTCACCTACCATATTTGTAACTGGTTCAGAAATCTCTAGCTCGATGTTGCTG 1007
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrrSerGlyLeuGluAla 258
Db 1008 AATGACTTCCTTAATGACCAAGATAGAGATAAAATAATATCTATAGTGGCCTGGAAGCC 1067
QY 259 PheGlyAspValLys 263
Db 1068 TTTGGGAGTGTWAAA 1082

RESULT 9
AY416671 1420 bp DNA linear GSS 17-DEC-2003

LOCUS
DEFINITION
Pan troglodytes HMGC52 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pan troglodytes (chimpanzee)
Pan troglodytes

REFERENCE
AUTHORS
1 (bases 1 to 1420)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Feriexera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)

JOURNAL
PUBMED
REFERENCE
AUTHORS
2 (bases 1 to 1420)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Feriexera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.

TITLE
JOURNAL
COMMENT
FEATURES
source
1..1420
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1420
/gene="HMGC52"
/locus_tag="HCM5968"

gene

ORIGIN

Alignment Scores:
Pred. No.: 2.17e-136 Length: 1420
Score: 1285.00 Matches: 243
Percent Similarity: 72.55% Conservative: 61
Best Local Similarity: 58.00% Mismatches: 73
Query Match: 51.17% Indels: 42
DB: 29 Gaps: 1

```
US-10-622-516-2 (1-478) x AY416671 (1-1420)
QY      17 ValGlyIleValAlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeu 36
      |||||:::|||||:::|||||
Db      160 GTGGGCATCCTGNNNCTGGAGGCTACTTCCNNNNNNNNNNNNNNNNNNNNNTGACTTG 219
QY      37 GluLysTyrAspGlyValAlaAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMet 56
      |||||:::|||||:::|||||
Db      220 GAGAAGTATTAACAATGTGGAGCAGGAAAGTATACAGTGGGCTTGGGCCAGACCCATATG 279
QY      57 GlyPheCysThrAspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeu 76
      |||||:::|||||:::|||||
Db      280 GGCTTCTGCTCAGTCCAAAGAGCATCAACTCCCTGTGCTGANNNTGGTCAACGGCTG 339
QY      77 MetGluArgAsnAsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThr 96
      |||||:::|||||:::|||||
Db      340 ATGGAGCGCATACAGCTCCCATGGGACTCTGTGGCAGGCTGGAAGTAGGCACCTGAGACC 399
QY      97 IleIleAspLysSerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGly 116
      |||||:::|||||:::|||||
Db      400 ATCATGTACAAGTCCAAAGCTGTCAAAAACAGTGCTCATGGAACCTTCCAGGANNCAGGC 459
QY      117 AsnThrAspIleGluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaVal 136
      |||||:::|||||:::|||||
Db      460 AATACTGATATTGAGGGCATAGATACCAACAATGCTGTCTACGCTGACTGCCTCCCTC 519
QY      137 PheAsnAlaValAsnTrpIleGluSerSerSerTrpAsp----- 149
      |||||:::|||||:::|||||
Db      520 TTCAATGCTGCCAACTGATGAGATCCAGTTCCTGGATGTCGTATATGCCATGTGTC 579
QY      149 ----- 149
Db      580 TGTGGAGACATTGCCGTCTATCCAGTGTATATGCTCGTCCCAAGTGGGCGGAGCT 639
QY      150 -----GlyLeuArgGlyThr 154
      |||||:::|||||:::|||||
Db      640 GTGGCTATGCTGATTTGGGCCCAAGCCCCCTCTGGCCCTGGAGCGAGGGCTGAGGGGAACC 699
QY      155 HisMetGlnHisAlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleVal 174
      |||||:::|||||:::|||||
Db      700 CATATGGAGAATGCGTATGACTTCTTCAAAACAAATTGTGGCTCGGAGTACCAATATATG 759
QY      175 AspGlyLysLeuSerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyr 194
      |||||:::|||||:::|||||
Db      760 GATGGGAAGCTTTCCATCCAGTGTACTTGCGGCTTGATCGATGTTACACATCATATAC 819
QY      195 CysLysLysIleHisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsn 214
      |||||:::|||||:::|||||
Db      820 CGTAATAAAATCCAGATCAGTGAAGCAAGCTGGCAGCGCATCGACCCCTTACCCCTTGAC 879
QY      215 AspPheGlyPheMetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAla 234
      |||||:::|||||:::|||||
Db      880 GATTTCAGTACATGATCTTTCATACACCCCTTTTGCAAGATGTCACAGAAGTCTTGCT 939
QY      235 ArgMetLeuLeuAsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSer 254
      |||||:::|||||:::|||||
Db      940 CGCTGATGTTCAATGACTTCTGTCAAGCCAGACAGTACACAAACCAAGCTTATATAG 999
QY      255 GlyLeuGluAlaPheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGlu 274
      |||||:::|||||:::|||||
Db      1000 GGGCTGAGGCTTTCGGGGGCTAAAGCTGGAAGACACCTACCAACAAGAGCACTGAT 1059
QY      275 LysAlaPheMetLysAlaSerSerGlnLeuPheSerGlnLysThrLysAlaSerLeuLeu 294
      |||||:::|||||:::|||||
Db      1060 AAAGCACTTCTAAAGGCTCTCAGGACATGTTGCACAGAAACCAAGGCTTCCCTTTAC 1119
QY      295 ValSerAsnGlnAsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerVal 314
      |||||:::|||||:::|||||
Db      1120 CTCTCCACGCACAATGGGAACATGTACACCTCATCCCTGTACGGGTGCTGCGCTT 1179
QY      315 LeuAlaGlnTyrSerProGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGly 334
      |||||:::|||||:::|||||
```

```
Db      1180 CTGTCCCACTCTGCCCCAGAACTGGCTGGCTCCAGATTGGTCCCTTCTTATGGC 1239
QY      335 SerGlyLeuAlaAlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySer 354
      |||||:::|||||:::|||||
Db      1240 TGTGTTTAGCAGCAAGTTCTTTTCATTTCAGATATCCAGATGCTGCTCCAGGCTCT 1299
QY      355 AlaLeuAspLysIleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThr 374
      |||||:::|||||:::|||||
Db      1300 CCCCTGACAAGTTGTGTCTCAGTACATCAGACCTGCCAAACGCTTAGCCTCCCGAAG 1359
QY      375 GlyValAlaProAspValPheAlaGluAsnMetLysLeuArgGluAspThrHis 393
      |||||:::|||||:::|||||
Db      1360 TGTGTGCTCTCAGAGAGTTCACAGAAATATGAATCAAGAAGACATTTACCAT 1416

RESULT 10
BM474567      998 bp      mRNA      linear      EST 05-FEB-2002
LOCUS      BM474567
DEFINITION      AGENCOURT_6489954 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5521787
5', mRNA sequence.
ACCESSION      BM474567
VERSION      BM474567.1 GI:18523609
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 998)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
      Email: cgabs-r@mail.nih.gov
      Tissue Procurement: ATCC
      cDNA Library Preparation: Life Technologies, Inc.
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
      DNA Sequencing by: Agencourt Bioscience Corporation
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LNL at:
      http://image.llnl.gov
      Plate: LLM12187 row: p column: 12
      High quality sequence stop: 759.
FEATURES
     source      1..998
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:5521787"
                  /tissue_type="leiomyosarcoma"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_lib="NIH_MGC_71"
                  /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                  Site_2: SalI; cloned unidirectionally. Primer: Oligo dt.
                  Average insert size 2.1 Kb. "
ORIGIN
Alignment Scores:
Pred. No.:      4.12e-134      Length:      998
Score:          1263.00      Matches:      250
Percent Similarity:      83.33%      Conservative:      0
Best Local Similarity:      83.33%      Mismatches:      4
Query Match:      50.30%      Indels:      46
DB:             12      Gaps:      1
US-10-622-516-2 (1-478) x BM474567 (1-998)
QY      1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
      |||||:::|||||:::|||||
Db      30 ATGCTTGATCACTTCTTGAATGCAAGAAGCTTGCTGGCCAAAAGATGTGGAATTGTT 89
QY      21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
      |||||:::|||||:::|||||
Db      90 GCCCTTGAGATATTTCTTCTCTCAATATGTTGATCAAGCAGAGTTGAAAAATATGAT 149
```

```
QY      41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
      |||
Db      150 GGCTAGATGCTGGAAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGGGCTTCTGCACA 209
QY      61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
      |||
Db      210 GATAGAGAAGATATTAATCTCTTTCATGACTGTGTTCAAGATCTTATGGAGAGAAAT 269
QY      81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
      |||
Db      270 AACCTTCTCTATGATTCGATTTGGGGGCTGGAAGTTGGAACAGACATCATCGACAAA 329
QY      101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
      |||
Db      330 TCAAACTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA 389
QY      121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
      |||
Db      390 GAAGGAATCGACACAACATAATGCATGCTATGAGGACACAGCTGCTGCTTCAATGCTGTT 449
QY      141 AsnTrpIleGluSerSerSerTrpAsp----- 149
      |||
Db      450 AACTGGATTGAGTCCAGCTCTGGGATGACCGGTATGCCCTGGTAGTTGCAGAGATATT 509
QY      149 ----- 149
Db      510 GCTGTATATGCCACAGAAATGCTTAGACCTACAGGTGAGTTGAGCAGATAGCTTGCTA 569
QY      150 -----GlyLeuArgGlyThrHisMetGlnHis 158
      |||
Db      570 ATGGGCCCAATGCTCCTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCCAAT 629
QY      159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
      |||
Db      630 GCCTATGATTTTACAAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAATAACTC 689
QY      179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
      |||
Db      690 TCCATACAGTGTACTCACTCAGTGCATTAAGACCGCTGCTATCTGCTACTGCAAAAAGATC 749
QY      199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
      |||
Db      750 CATGCCCAGTGGCAGAAAAGAGGAATGATNAGATTTTACCTGGAATGATTTGGCTTC 809
QY      219 MetIlePheHisSerProTyr-CysLysLeuValGlnLysSer-LeuAlaArgMetLeuL 238
      |||
Db      810 ATGATCTTTCACCTCACATATTTGTAACCTGTTCAGAATACTNCTAGCTCGATGTTGC 869
QY      238 euAsnAspPhe-LeuAsnAspGlnAsnArgAspLysAsn-SerIleTyrSer 254
      |||
Db      870 TGAATGACTTCCNTTAATGACCNGAATAGAGATAAAATAGNTATCTATAGT 921
RESULT 11
BX419944      1201 bp      mRNA      linear      EST 13-MAY-2003
LOCUS      BX419944 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION      CS0DF022YE18 5-PRIME, mRNA sequence.
ACCESSION      BX419944
VERSION      BX419944.1 GI:30650862
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 1201)
AUTHORS      Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
```

Invitrogen. This sequence belongs to sequence cluster 3098.r For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DF022BC09QP1&cluster=3098.r>. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DF022BC09QP1.

FEATURES

source

```
1. 1201
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="CS0DF022YE18"
   /tissue_type="FETAL BRAIN"
   /dev_stage="fetal"
   /clone_id="Homo sapiens FETAL BRAIN"
   /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
```

ORIGIN

Alignment Scores:

Pred. No.:	5.61e-134	Length:	1201
Score:	1263.00	Matches:	256
Percent Similarity:	81.90%	Conservative:	2
Best Local Similarity:	81.27%	Mismatches:	12
Query Match:	50.30%	Indels:	48
DB:	13	Gaps:	3

US-10-622-516-2 (1-478) x BX419944 (1-1201)

```
QY      1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
      |||
Db      90 ATGCTGGATCATTCTCTTGAATGCAGAAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 149
QY      21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
      |||
Db      150 GCCCTTGAGATCTATTCTCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 209
QY      41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
      |||
Db      210 GGTGTAGATGCTGAAAGTATACCATTTGGCTTGGGCCAGGCCAAGATGGGCTTCTGCACA 269
QY      61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
      |||
Db      270 GATAGAGAAGATATTAATCTCTTTCATGACTGTGTTCAGAATCTTATGAGAGAGAAAT 329
QY      81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
      |||
Db      330 AACCTTCTCTATGATTCGATTTGGGGGCTGGAAGTTGGAACAGACAAATCATCGACAAA 389
QY      101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
      |||
Db      390 TCAAACTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA 449
QY      121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
      |||
Db      450 GAAGGAATCGACACAACATAATGCATGCTATGAGGACACAGCTGCTGCTTCAATGCTGTT 509
QY      141 AsnTrpIleGluSerSerSerTrpAsp----- 149
      |||
Db      510 AACTGGATTGAGTCCAGCTCTTGGGATGACCGGTATGCCCTGGTAGTTGCAGAGATATT 569
QY      149 ----- 149
Db      570 GCTGTATATGCCACAGAAATGCTTAGACCTACAGGTGAGTTGGAGCAGTAGCTTGCTA 629
QY      150 -----GlyLeuArgGlyThrHisMetGlnHis 158
      |||
Db      630 ATGGGCCCAATGCTCCTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT 689
```


QY 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
|||
Db 690 GCCTATGATTTTACAGCCCTGATATGCTATCTGAATATCCTATAGTAGATGGAAAACTC 749

QY 179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
|||
Db 750 TCCATACAGTGCTACTCACTCAGTCATTAGACCCTGTCTACTCTGCTACTGCCAAAAAGATC 809

QY 199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAsp-PheThrLeuAsnAspPheGlyPh 218
|||
Db 810 CATGCCCAGTGGCAGAAGAAGAAATGATAAAGATTTTTTAACCTTGAATGATTTTGGCTT 869

QY 218 emetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeu 238
|||
Db 870 CATGATCTTTCACTCACCATATTTGTAACACTGGTTCAAGAAATCTTAAGCTCGGATGTGCT 929

QY 238 uAsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAl 258
|||
Db 930 GAATGACTCTCTTAATG---MCAGATAGAGATTA-AAATAGTAC-TATAGTGGCCTGGAGCC 984

QY 258 apHeGlyAspValLysLeuGluAspThrTyrPheAspArgAsp 272
|||
Db 985 -TTTGGGAGKGTAAATWAGARGCAC--CTCTTTGATAGAGAG 1023

LOCUS	BU508388	905 bp	mRNA	linear	EST 12-SEP-2002
DEFINITION	AGENCOURT_10099759	NIH_MGC_71	Homo sapiens	cdna clone	IMAGE:65024655
	5', mRNA sequence.				

ACCESSION	BU508388
VERSION	BU508388.1
	GI:22814621

SOURCE ORGANISM	Homo sapiens (human)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

REFERENCE	1 (bases 1 to 905)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: L14058 row: j column: 02
High quality sequence stop: 717.

FEATURES	Location/Qualifiers
Source	1. .905

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6502465"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 71"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

```

ORIGIN

Alignment Scores:	2.56e-133	Length:	905
Pred. No.:	1255.50	Matches:	247
Score:	82.94%	Conservative:	1
Percent Similarity:	82.61%	Mismatches:	6
Best Local Similarity:	50.00%	Indels:	45
Query Match:	13	Gaps:	2
DB:			

US-10-622-516-2 (1-478) x BU508388 (1-905)

9 AlaGluAlaCysTrpProLysAspValGlyIleValAlaLeuGluIleTyrPheProSer 28

Db 12 GCAGAGCTTGCTGGCCAAAGATGTGGGAATTGTGCCCTTGAGATCTATTTCCCTCT 71

QY 29 GlnTyrValAspGlnAlaGluLeuGluLysTyrAspGlyValAsp-AlaGlyLysTyrTh 48

Db 72 CAATATGTCATCAGCAGCTTGAAATAATGATGCTAGATGCGTGAAGTATAC 13

48 rIleGlyLeuGlyGlnAlaIysMetGlyPheCysThrAspArgGluAspIleAsnSerLe 68

Db 132 CATGGCTGGCCAGCCAAGATGGGCTTCTGCACAGATAGAGAAGATATTACTCTCT 15

68 ucysmethrvalvalglmasnleumetgluargasnleusertryasp cystllecl 888

Db 192 TTGCATGACTGTGTT CAGAATCTTATGGAGAGAAATAACCTTTCCTATGATTGCATTGG 255

88 yargleuGuaIcIyThrgIuthrIeIleasplySerlysserVallysthrasne 10

Db 252 GCGGCTGGAAGTTGGAACAGAGACAATCATCGACAATCAAAGTCTGTGAAGACTAATT 31

QY 108 umetGlnLeupheGluGluSerGlyAsnThrAspIleGluGlyIleAspThrThrAsnAl 12

Db 312 GATGCAGCTGTTGAAGAGTCTGGGAAACAGATATAGAAGGAATCGACACCACTTAATGC 3

QY 128 acysttyrglgl ythr aa,ava, pneasna lavalas ntrpl legluser ser ser ltr 14

D6 372 ATGCTATGGAAGCACAAGCTGCTGCTTCATAGCTTGTAACUGGATTCAGTCCAGCTCTTG 433

QY 148 PASP

DB 432 GGATGGACGGTAAGCCCTGGIAGTGCAGGAGATAITGCIGATAAGCCACAGGHAATC 433

149 ----- 149

DD #92 1A0HCL1ACABUIGAHU110GABUCAB1ABU1C10C1FH110G0CCFHH10C1C11AFHH1

100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040

[illegible]

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

CA512392
DOCTIS
CA512392
200 hr
PBNM 140000
EST 15-NOV-30

DEFINITION	UI-R-FJ0-cpz-j-19-0-UI.r1	UI-R-FJ0	Rattus norvegicus	CDNA clone
UI-R-FJ0	UI-R-FJ0	UI-R-FJ0	Rattus norvegicus	CDNA clone

ACCESSION	CA512392
VERSION	CA512392 1
	GT:25003345

KEYWORDS EST. **batius norvegicus (Norway rat)**

ORGANISM Rattus norvegicus
Eukaryota: Chordata: Vertebrata: Euteleostomi:

REFERENCE 1 (bases 1 to 789)
AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9) , 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) .
Seq primer: M13 REVERSE.

FEATURES
source
1. .789
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-FJ0-cpz-j-19-0-UI"
/issue_type="embryo"
/dev_stage="embryo"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-FJ0"
/note="Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I;
UI-R-FJ0 is a cDNA library containing the following
tissue(s): rat embryo. The library was constructed
according to Ronaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CACTCTACT. This library
was created for the University of Iowa Program for Rat
Gene Discovery and Mapping (Val Sheffield, Bento Soares
and Tom Casavant)"

ORIGIN
Alignment Scores:
Pred. No.: 1.94e-132 Length: 789
Score: 1247.00 Matches: 244
Percent Similarity: 94.66% Conservative: 4
Best local Similarity: 93.13% Mismatches: 14
Query Match: 49.66% Indels: 0
DB: 14 Gaps: 0

US-10-622-516-2 (1-478) x CA512392 (1-789)
QY 190 CysTyrSerValTyrCysLysLysIleHisAlaGlnTrpGlnLysGluGlyAsnAspLys 209
Db 3 TGCTATTCTGTCTACCGCAAAAGATCCGGGCCAGTGGCAGAAAGAGGAAAGCATAA 62
QY 210 AspPheThrLeuAsnAspPheGlyPheMetIlePheHisSerProTyrCysLysLeuVal 229
Db 63 GATTTTACCCTGAATGATTTGGCTTCATGATCTTTCACTCCGCATACTGTAAACTGGTG 122
QY 230 GlnLysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsnAspGlnAsnArgAspLys 249
Db 123 CAGAAATCTCTAGCTAGATGTCTGATGACTTTCTTAACGATCAAAACAGACAA 182

QY 250 AsnSerIleTyrSerGlyLeuGluAlaPheGlyAspValLysLeuGluAspThrTyrPhe 269
Db 183 AACAGTATTTACAGTGGGCTGGAAAGCCTTTGGGGATGTGAATTAGAAGATACTACTTC 242
QY 270 AspArgAspValGluLysAlaPheMetLysAlaSerSerGluLeuPheSerGlnLysThr 289
Db 243 GACAGAGATGTGAAAGGCATTATTGAAGGCTAGTGTGAGCTATTTCACCAAGAAACA 302
QY 290 LysAlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyrThrSerSerValTyrGly 309
Db 303 AAGGCATCTTGTCTGTATCGAATCAAAATGGAACATGTACACATCCTGTATACGGT 362
QY 310 SerLeuAlaSerValLeuAlaGlnTyrSerProGlnGlnLeuAlaGlyLysArgIleGly 329
Db 363 TCCCTTGCTTCTGTCTGGCACAGTACTCACTCAACAGTTGGCCGGGAAGAGATTGGA 422
QY 330 ValPheSerTyrGlySerGlyLeuAlaAlaThrLeuTyrSerLeuLysValThrGlnAsp 349
Db 423 GTGTTCTCTTACGGTTCTGGCTTGCTGGCCACACACTTACTCCCTTAAAGTCACAGAAT 482
QY 350 AlaThrProGlySerAlaLeuAspLysIleThrAlaSerLeuCysAspLeuLysSerArg 369
Db 483 GCCACACCAAGATCTGCTCTTGACAAATAACAGCAAGTTTATGTGACCTTAAGTCAAAG 542
QY 370 LeuAspSerArgThrGlyValAlaProAspValPheAlaGluAsnMetLysLeuArgGlu 389
Db 543 CTTGACTCAAGACGTGTGTGGCACACAGCGTCTTGCTGAATAATGAAGCTCAGAGAG 602
QY 390 AspThrHisLysLeuValAsnTyrIleProGlnGlySerIleAspSerLeuPheGluGly 409
Db 603 GACACACATCACTTAGCCCACTATATTCCGCCAGTGTCAATAGATTCACTCTTCGAAGGA 662
QY 410 ThrTrpTyrLeuValaArgValaAspGluLysHisArgArgThrTyrAlaArgArgProThr 429
Db 663 ACATGGTATCTAGTCAGAGTGGATGAAGACACAGAAGAACTTACGCCCGCTCCTCC 722
QY 430 ProAsnAspAspThrLeuAspGluGlyValGlyLeuValHisSerAsnIleAlaThrGlu 449
Db 723 ACANNATGACACAGTTTGATGAAGAGAGTGGAGCTTGTGATTGANNACACAGCTACAGAG 782
QY 450 HisIle 451
Db 783 CATATT 788

RESULT 14
CD354310 779 bp mRNA linear EST 15-JUL-2003
LOCUS UI-M-GMO-cgc-j-20-0-UI.r1 NIH_BMAP_GMO Mus musculus cDNA clone
DEFINITION IMAGE:30360907 5', mRNA sequence.
ACCESSION CD354310 GI:31146811
VERSION CD354310
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS 1 (bases 1 to 779)
TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgarbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousef1.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. .779

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30360907"
/tissue_type="whole brain"
/dev_stage="1, 5 and 15 days newborn"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP GM0"
/note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according Bonaïdo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGACTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

Alignment Scores:
Pred. No.: 7.04e-129 Length: 779
Score: 1216.00 Matches: 239
Percent Similarity: 94.59% Conservative: 6
Best Local Similarity: 92.28% Mismatches: 14
Query Match: 48.43% Indels: 0
DB: 14 Gaps: 0

US-10-622-516-2 (1-478) x CD354310 (1-779)

QY 219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeu 238
Db 3 ATGATCTTCACTCACCATATGTGAACCTGGTGAGAAATCTTAGCTCGATGTTCTTG 62
QY 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db 63 AATGACTTCTTAATGATCAAAACAGAGATAAAAACAGTATTACAGTGGACTGGAAGCC 122
QY 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db 123 TTTGGGGACGTTAATTAGAAGTACTTACTTTGACAGAGATGTAGAAAAAGCATTTATG 182
QY 279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db 183 AAGCTAGTTCTGAGCTATTCAACAGAAACAAAGGCGCTTTGCTTGTGTTAATCAG 242
QY 299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db 243 AATGAAATATGTACACATCCTGTCGTATGTTCCCTGGCTTCTGCTGGCACAGTAC 302
QY 319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db 303 TCACCTCAGCAGTTGGCAGGAGAGGGTTGGAGTGTCTTACGGTTCTGGCTTGCT 362
QY 339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db 363 GCCACGCTGTACTCCCTTAAAGTACACAGAAGCCACACCGAGATCTGCCCTTGATAA 422
QY 359 IleThrAlaSerLeuCyAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db 423 ATTAACAGCAAGTTTATGTGACCTTAAATCAAGGCTTGACTCGAGAACTTGTGTGGCACCG 482
QY 379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db 483 GATGTCTTGTGAAACATGAGCTCAGAGAGACACACATCACTTAGCCAACTATATT 542
QY 399 ProGlnGlySerIleAspSerLeuPheGlnGlyThrTyrTyrLeuValArgValAspGlu 418

Db 543 CCCAGTGTTCATAGACTCACTCTTTGAAGGAACGTGATCTGTGTCAGAGTGATGAA 602
QY 419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
Db 603 AAACACAGAAAGACTTACCGCGCGCCCTTCACANATGACCAAGTTGGATGAAGGA 662
QY 439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
Db 663 ATGGGGCTCGTGATAGTAAACAGCAACAGACATATTCAGCCCTGCTAAGANAGT 722
QY 459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGlu 477
Db 723 GCAAGACTCCCTGCAACCTCGGNCGAATCTGAATCAGCTGCATCAGTAAACGGGAG 779

RESULT 15
BU149049
LOCUS
DEFINITION
AGENCOURT_8743315 NIH_MGC_129 Mus musculus cDNA clone IMAGE:63922260 5', mRNA sequence.
ACCESSION
BU149049
VERSION
BU149049.1 GI:22662581
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM1381 row: n column: 05
High quality sequence stop: 599.
location/Qualifiers

FEATURES

source

1..967
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:63922260"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_129"
/note="Organ: Olfactory epithelium; Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC library."

ORIGIN

Alignment Scores:
Pred. No.: 4.95e-128 Length: 967
Score: 1210.00 Matches: 249
Percent Similarity: 80.82% Conservative: 8
Best Local Similarity: 78.30% Mismatches: 16
Query Match: 48.19% Indels: 47
DB: 13 Gaps: 1

US-10-622-516-2 (1-478) x BU149049 (1-967)

QY 3 GlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleValAlaLeu 22
Db 22 GGGTCACCTTCCTTTGAATGCAGAGGCTTGCTGGCCAAAGATGTGGAAATCGTGCCTT 81
QY 23 GlnIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAspGlyVal 42

Db 82 GAAATCTACTTTCCTTCTCAATATGTCGATCAAGCTGAGTTGGAATAATACGATGCTGTA 141

QY 43 AspAlaGlyLysThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThrAspArg 62
|||||
Db 142 GATGCTGGAAGATATACCATCGGCTGGGCCAGGCCAGGATGGGCTTCTGCACGGATCGT 201
|||||

QY 63 GluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsnAsnLeu 82
|||||
Db 202 GAAGACATCAACTCTCTTTGCTGCTGACTGTGTTCAAGAACTGATGGAGAGACATAGCCCTT 261
|||||

QY 83 SerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLysSerLys 102
|||||
Db 262 TCCTATGATTGCATTGGGCGGCTAGAGTTGGAACAGACAAATCATCGACAAATCGAAA 321
|||||

QY 103 SerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIleGluGly 122
|||||
Db 322 TCAGTGAAGTCTTAATTGATGACAGCTGTTTGAAGAGTCTGGGAATACAGATATAGAAAGA 381
|||||

QY 123 IleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaValAsnTrp 142
|||||
Db 382 ATAGATACAACCAATGTCATGCTATGGGGGCACAGCTGCAGTCTTCAATGCCGTGAAGCTGG 441
|||||

QY 143 IleGluSerSerSerTrpAspGly----- 150
:::|||||

Db 442 GTCGAATCCAGCTCTGGGATGACGATATGCTCTGTAGTTGACAGACATGCTATA 501
150 ----- 150

QY 502 TATGCCACAGAAATGCCAGACTACAGGTGAGTTGAGCTGTGGCCCTGCTAATTGGG 561

Db 151 -----LeuArgGlyThrHisMetGlnHisAlaTyr 160
|||||

QY 562 CCAACGCTCCTCTAATTTTGTACCGANGGCTCCGTGNACACACATGCAGCATGCCCTAT 621

Db 161 AspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIle 180
|||||
Db 622 GACTTTTACAAAGCTGACATGCTCTCCAGTACCTGTGTGTCGACGAAAGCTTCATATA 681
|||||

QY 181 GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAla 200
|||||
Db 682 CAGTGCTAACCTCAGCGGCCCTGGACCGCTGCTATTCTGTCTACCGCAAAAAGATCCGTGCC 741
|||||

QY 201 GlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIle 220
|||||
Db 742 CAGTGGCAGAAAGAGGAAAGATAAAGA-TTTACCCCTGAATGATNTTGGCTTCATGATC 800
|||||

QY 221 PheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp 240
|||||
Db 801 TTTCACTCACCATATTGTAACTGTGCAGAAATCTCTAGCTCGATGTCCCTGAATGAC 860
|||||

QY 241 PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla-PheGly 260
|||||

Db 861 TTNTCCTATGATCAAAACAGAGATAA-AAACAGTATTACAGTGAGACTGGAAGCCTTTTGG 919
|||||

QY 260 YAsp-ValLysLeuGluAspThrTyr-PheAspArgAspValGluLys 275
|||||
Db 920 GGACGNNTAATTAGAAAGATACTTTTGACCGAGATGTAAAAAAG 967
|||||

Search completed: June 24, 2004, 15:57:53
Job time : 2886 secs